Journal of Experimental Psychology

Volume 91, 1971

PUBLISHED MONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA. 17604
AND 1200 SEVENTEENTH ST. N. W., WASHINGTON, D. C. 20036

Second-class postage paid at Lancaster, Pa.

CONTENTS OF VOLUME 91

Special Announcement

Editing this Journal is very much a communal enterprise. Each number of the Journal carries on its cover the names of the gifted, conscientious Consulting Editors whose generous contributions of talent, time, and energy are of inestimable value to the Action Editors in arriving at editorial decisions and to the authors in revising their manuscripts or even their research strategies so as to contribute more effectively to the science of psychology. Aside from the gratitude of the Action Editors and many of the authors, the only rewards the Consulting Editors receive for their contribution to scientific communication and the postdoctoral training of some of the contributing authors are their own personal satisfaction in doing a splendid job, a complimentary subscription to the Journal, and such prestige and other consequences that flow from having their names listed on the cover in what is obviously very good company.

Each year a large number of manuscripts are referred to reviewers who are not among those listed on the masthead. These unlisted reviewers possess the same admirable characteristics as the regular Consulting Editors, and they give us the same kind of thorough and competent reviews as do the latter. The unlisted reviewers receive no tangible rewards for their contributions; their names do not appear on the masthead; they receive no complimentary subscription. They deserve more recognition. Over fifty of the manuscripts received during 1970 were reviewed by over forty-six such unlisted reviewers.

Some of these reviewers have joined the regulars, so that their names grace the 1971 mastheads. Most of them, however, did not, and we take this opportunity to recognize more explicitly their contributions to the authors and editors of the Journal by listing their names below:

Harry P. Bahrick
Richard S. Bogartz
Wilfred J. Brogden
Robert B. Cairns
E. John Capaldi
Charles N. Cofer
James H. Davis
Robyn M. Dawes
Wesley M. DuCharme
Howard E. Egeth
William Epstein
James R. Erickson
Morton P. Friedman
Wendell R. Garner
Anthony G. Greenwald
Kenneth R. Hammond
Ray Hyman
Lawrence E. Jones
William E. Kappauf
Howard H. Kendler
Walter Kintsch
Marvin Levine
Dominic W. Massaro
Dale W. McAdam

Raymond C. Miles
Robert E. Morin
Clyde E. Noble
Allan U. Paivio
Allan J. Pantle
Cameron R. Peterson
Gordon F. Pitz
Irwin Pollack
David C. Raskin
J. Edward Russo
Peter H. Schiller
Marian S. Schwartz
Norman J. Slamecka
Paul Slovic
Herbert H. Stenson
Willard R. Thurlow
Constantine Trahiotis
Benton J. Underwood
Paul C. Vitz
Daniel Weintraub
Nancy Wiggins
Keith A. Wollen
John I. Yellot

The Journal is indebted to them.

David A. Grant, *Editor*

Journal of
Experimental Psychology

Vol. 91, No. 1 November 1971

CHANGE IN CONVERGENCE AND RETINAL DISPARITIES AS AN EXPLANATION FOR THE WALLPAPER PHENOMENON [1]

HIROSHI ONO, LANCE MITSON, and KAREN SEABROOK [2]

York University

A. Ames' notion that the shift in apparent distance observed in the wallpaper phenomenon is due to a change in retinal disparities rather than a change in convergence was investigated. In Exp. I, repeating patterns on a Veith-Mueller circle and on a frontal parallel plane were presented to 18 *S*s. Apparent distance decreased as convergence level increased in both situations, supporting the notion that convergence as a cue is involved. In Exp. II, a repeating pattern was presented to 30 *S*s such that convergence level was constant but disparities varied. Apparent distance was a function of disparities only when *S*s became familiar with the pattern, indicating that known size of the pattern and disparities together serve as a cue but not disparities per se. An argument is made that the explanation for the wallpaper phenomenon does not rest on a single cue and that the apparent distances obtained were an outcome of the visual system integrating different and conflicting sets of information.

When a uniform repeating pattern is viewed with certain vergences so that pairs of elements in the pattern are fused, the entire pattern changes in apparent distance. This effect, called the "wallpaper" phenomenon, has been explained in two different ways. The traditional explanation is that convergence serves as a cue for determining the apparent distance (Helmholtz, 1925). In support of this explanation, a recent experiment by Lie (1965) showed nearly perfect correspondence between apparent distance and convergence distance. The second explanation is that the shift in apparent distance is due to the changes in retinal disparities of the pairs of elements that accompany the change in convergence (Ittelson, 1960). Disparity in this situation refers not to the degree of noncorrespondence produced by a single point, but to that produced by two corresponding elements of the pattern which are fused. A geometric derivation with a pictorial illustration of the change in disparities as a function of vergence appears in Ittelson (1960, pp. 123–127). In support of this explanation, Ames (described in Ittelson, 1960) demonstrated a lack of apparent-distance shift when a repeating pattern arranged in a Vieth-Mueller circle was viewed with different vergences and also a shift in apparent distance when vergence was kept constant

[1] This research was supported by Grant A0296 from the National Research Council of Canada.

[2] Requests for reprints should be sent to Hiroshi Ono, Department of Psychology, York University, Toronto, Ontario, Canada.

but disparities were changed by the use of prisms.

The second explanation can be criticized on a logical ground (Ono, 1970). A difficulty arises from the fact that retinal disparities per se can provide information for exocentric distance, but not for egocentric distance (cf. Gogel, 1971; Ogle, 1962). When one of the points is *O* and the other is some external point, the extent is labeled absolute or egocentric distance. When the two points are both external, the extent between the points is called depth or exocentric distance. The wallpaper phenomenon is a misjudgment of an egocentric distance; and, therefore, the difference in disparities provided by changes in convergence cannot be an explanation. The present argument implies that changes in disparities may be involved in Ames' results, but *S* must be using other information or cues instead of, or as well as, the disparities to make the judgment.

The concern of the two experiments reported here is with the two experimental settings studied by Ames. The investigations made by Ames were more in the nature of demonstrations than experiments. The present experiments were conducted in a more controlled setting. Experiment I dealt with measurements of apparent distance of a repeating pattern on a Vieth-Mueller circle and on a frontal parallel plane for different levels of vergence. Experiment II dealt with measurements of apparent distance of a repeating pattern on a frontal parallel plane under two viewing conditions. The two conditions differed in disparities, but vergence was kept constant by using wedge prisms.

Experiment I

The apparent distance of a repeating pattern on a Vieth-Mueller circle and on a frontal parallel plane was measured at four levels of vergence. The underlying notion was that in both viewing situations there are two sets of opposing information. The apparent distance presumably reflects the way the visual system resolves the conflict. Viewing a repeating pattern on a frontal parallel plane, the visual system must cope with competing information. Some information, such as the knowledge of the actual location, accommodation, known size plus constant retinal image size, and perhaps other factors, indicates that the screen has not moved. Opposing information such as that from convergence indicates that the screen has moved closer to *O*. Additional opposing information could be the known screen size plus increased disparities. Because disparities produced by a given object are correlated with egocentric distance, a change in disparities can be a cue to distance, if the size of the object is known or assumed to be equal at different distances. Presumably, the visual system processes both sets of information and reaches a compromise. In the case of viewing the pattern on a Vieth-Mueller circle, changes in convergence produce (theoretically) no change in disparities. Hence, there is no information other than convergence to compete with the information indicating no screen movement. The hypothesis tested was that there is a wallpaper effect in both viewing situations, but that the agreement between the vergence levels and the apparent distances should be greater for the pattern presented on the frontal parallel plane.

Method

Apparatus.—Three apparatuses with wire mesh screens (mesh size 1.3 cm.²) were built. One apparatus had a flat screen 41 cm. high × 46 cm. wide, mounted between 51 × 92 cm. plywood plates. The screen subtended approximately 45° at a viewing distance of 51.8 cm. A second apparatus consisted of a cylinderical screen 51.8 cm. in diameter and 41 cm. high suspended between two plywood disks. A 20-cm.-wide strip was cut out to accommodate a headrest. A third apparatus consisted of a section of a cylinderical screen 51.8 cm. in diameter and 41 cm. high cut down to subtend approximately 45° from the bridge of the nose. The cylindrical section was mounted between plywood disks mounted on two vertical supports 20 cm. apart, one on each side of the headrest. This apparatus had approximately equal screen area to the flat screen of the first apparatus. Each apparatus was provided with a removable thin rod projecting vertically down into *S*'s field of view such that the tip was at eye level and straight ahead. The tip of the rod was used to indicate to *S* the appropriate vergence. In addition, one vertical wire forming a boundary of a single mesh in the median plane and

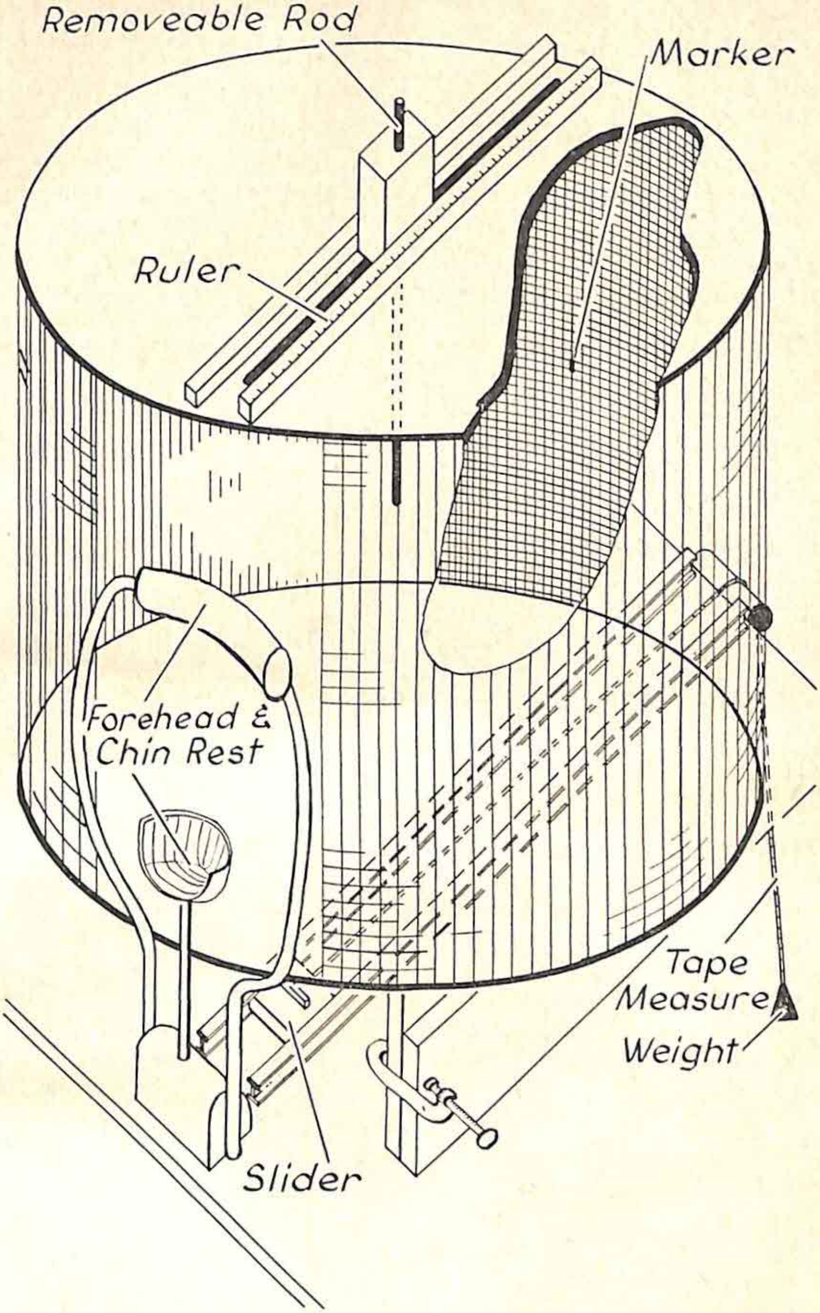

FIG. 1. Schematic drawing of one of the three apparatuses used in Exp. I.

at eye level was distinguished with black tape. The black marker enabled E to monitor S's convergence level. The second apparatus is illustrated in Fig. 1.

All the apparatuses were designed to clamp on a single base provided with a chin and forehead rest combination. Also mounted on the base was a sliding block which S used to indicate the apparent distance of the screen. The S's hand and the slider were covered by the bottom plate of the apparatus and thus no visual feedback was possible. Attached to the slider was a scale enabling E to read S's setting.

In addition to normal room lighting, a small fluorescent light placed behind and above S's head was used to light the screen. The screens were viewed against a white background.

Experimental design.—The three viewing conditions formed the major experimental variable. They

were: (*a*) Frontal Parallel Plane, (*b*) Complete Vieth-Mueller Circle, and (*c*) Partial Vieth-Mueller Circle conditions. The reason for the two different Vieth-Mueller conditions was to control for the visual extent of the repeating pattern. If there is no difference in apparent distance between Cond. *b* and *c* but *b* and *c* differ from *a*, then inferences about the differences in the apparent distance of repeating patterns on a Vieth-Mueller circle and a frontal parallel plane can be made. However, if Cond. *a* and *c* produce the same apparent distance which is different than that of *b*, the critical variable is the visual extent of the repeating pattern.

An experimental session consisted of presenting one of the three viewing conditions to *S* and obtaining measures of the apparent distances of the screen at different levels of convergence. All *Ss* served in all conditions. To control order effects, the sequence of the three viewing conditions was completely counterbalanced across 18 *Ss*. There were six sequences with three *Ss* in each sequence. In each session, four convergence distances were required of each *S*, corresponding to perceived separations of the diplopic images of the marker of 0, 1, 2, and 3 units (1 unit being defined as one square of the wire mesh). The apparent distance of the screen was measured four times for each of the four convergence distances. The convergence distances were grouped into four blocks such that all four convergence distances appeared in each of the blocks. The four blocks comprised a Latin-square design.

Procedure.—Before the first experimental session began, the interocular distance of each *S* was measured and the appropriate convergence distances for the three perceived separations of the marker were calculated. The chin rest was adjusted for each *S* so that the line of centers of *S*'s eyes was at the same height as the marker and so that the median plane of *S* was on the midline of the apparatus. The distance between the estimated nodal point of each eye and the screen was then adjusted with the aid of a guide rod so that the nodal point of each eye lay on the circumference of the Vieth-Mueller circles and an equal perpendicular distance away from the frontal parallel plane screen.

Each experimental session consisted of a practice and a test period. In the practice period, each *S* looked directly at the tip of the rod projecting down into his field of view and reported when the resulting diplopic images of the marker were stably separated. The rod was then withdrawn and *S* was asked to hold the images apart for 20 sec. This procedure was repeated until 4 successful attempts were made for each of the three required separations of the diplopic images of the marker. If *S* failed to achieve 4 successful attempts in a total of 10 tries, he was rejected from the experiment. Practice periods for the various convergence distances were separated by a rest period of 1 min.

Before the test period, *E* stressed to each *S* that the task was to discard any preconceived idea of the actual screen distance and to report only the perceived distance of the portion of the screen being directly regarded. The *S* grasped the handle of the slider with the thumb and forefinger of the preferred hand, and, when the rod was removed, set the slider beneath the apparent location of the screen. If any change in the image separations of the marker occurred before the response, the trial was disregarded and another run immediately. The *S* was asked to close his eyes between each trial. One-minute rest periods were given between each of the four blocks of trials.

Subjects.—A total of 41 men participated in the experiment. Of these, 18 met the criterion and ran in all three sessions. Sixteen *Ss* were from local high schools, and 2 were from the university community. All *Ss* had 20/20 vision in both eyes. The *Ss* were paid $2.00 per session.

Results and Discussion

In each viewing condition, the data obtained were four measurements of apparent distance for each of four convergence distances. A polynomial regression analysis was performed on the data for each *S* under each viewing condition. For each analysis, the total sum of squares was partitioned into linear, quadratic, and cubic sources. In the Frontal Parallel Plane condition, the linear source was largest for 16 out of 18 *Ss* and was statistically significant ($p < .05$) for all 16 *Ss*. The mean percent of variance over 18 *Ss* due to the linear source (100 × linear sum of squares/total sum of squares) was 62.8%, due to the quadratic source was 6.9%, and due to the cubic source was 1.5%. In the Complete Vieth-Mueller Circle condition, the linear source was largest for 14 out of 18 *Ss* and was statistically significant ($p < .05$) for all 14 *Ss*. The mean percent of variance due to the linear source was 43.8%, due to the quadratic source was 8.3%, and due to the cubic source was 7.4%. In the Partial Vieth-Mueller Circle condition, the linear source was largest for 15 out of 18 *Ss* and was statistically significant ($p < .05$) for 11 out of the 15 *Ss*. The mean percent of variance due to the linear source was 40.3%, due to the quadratic source was 8.9%, and due to the cubic source was 3.1%.

This outcome was considered to be sufficient justification for applying a simple linear regression analysis to the data for all *Ss*. For each *S*, the slope and the intercept of the regression line for the apparent distance on the convergence

distance were determined for each viewing condition. The means and the standard deviations of the slopes and intercepts are presented in Table 1. In the table, two sets of means and standard deviations are shown for each viewing condition; one set was obtained from all 18 *S*s, and the second set was obtained only from the first experimental session. The values in the second set are those without the effects of being exposed to the other viewing conditions. Each value in the second set was obtained from 6 *S*s, not 18.

Two-way (3 × 6) analyses of variance with correlated measures on one factor were performed on the two scores described above. The variables were the 3 viewing conditions and the 6 orders of presentation. The analyses showed that the main effects of viewing condition were statistically significant ($p < .001$) for the slope and intercept scores, $F\ (2, 24) = 13.53$ and $F\ (2, 24) = 16.20$, respectively. The assumption of equal covariance for a repeated-measurements design was not tested. However, the Geisser-Greenhouse univariate conservative F test still indicated significant differences among conditions at $p < .001$. The main effects in the slope and intercept scores were due to the Frontal Parallel Plane condition having a higher mean slope and lower mean intercept relative to the other two conditions. The Scheffé multiple comparison of slope scores showed that the Frontal Parallel Plane condition was significantly different from the Partial Vieth-Mueller Circle condition and from the Complete Vieth-Mueller Circle condition ($p < .01$), but the two Vieth-Mueller circle conditions were not significantly different from each other. The lack of difference between the two conditions indicates that the visual extent of the repeating pattern did not play a major role in the outcome of the experiment.

Also statistically significant were the interactions of Condition × Order for the slope and intercept scores, $F\ (10, 24) = 2.23$, $p < .06$, and $F\ (10, 24) = 2.82$, $p < .05$, respectively. The interactions were partly due to the fact that the viewing

TABLE 1

MEANS AND STANDARD DEVIATIONS OF SLOPES AND INTERCEPTS FROM THREE TEST SESSIONS COMBINED AND FROM FIRST TEST SESSION FOR THREE VIEWING CONDITIONS IN EXPERIMENT I

Measure	Cond.		
	Frontal Parallel	Partial V-M	Full V-M
All three sessions			
Slope			
$\bar{X}$	.52	.25	.27
SD	.25	.26	.22
Intercept			
$\bar{X}$	21.71	35.74	33.94
SD	14.56	15.19	12.49
First test session			
Slope			
$\bar{X}$	.68	.33	.32
SD	.19	.28	.08
Intercept			
$\bar{X}$	14.57	27.95	32.49
SD	11.32	13.40	5.03

condition tested first tended to give higher slope and lower intercept. This tendency can be seen in Table 1.

The results supported the hypothesis that the Frontal Parallel Plane condition should yield an apparent distance closer to the convergence distance than should be the case for the two Vieth-Mueller circle conditions. This conclusion was indicated by the steepest slope in the Frontal Parallel Plane condition. (Complete agreement would be a slope of unity). Also relevant to the hypothesis is the fact that the slopes in the two Vieth-Mueller conditions were greater than zero. A comparison of each mean against zero using t tests yielded $t\ (17) = 3.90$, $p < .002$ (two-tailed), for the Partial Vieth-Meuller Circle condition and $t\ (17) = 5.05$, $p < .002$ (two-tailed), for the Complete Vieth-Mueller Circle condition. The fact that the mean slopes in the two conditions were significantly greater than zero contradicts Ames' observation that the apparent distance remains constant for different convergence distances in the absence of change in disparities. The results of the present experiment are consistent with the traditional explanation that convergence as

a cue is involved with the wallpaper phenomenon.

Furthermore, the mean slope of .52 obtained in the Frontal Parallel Plane condition is not entirely consistent with Lie's (1965) findings showing almost perfect correspondence between the convergence distance and apparent distance, i.e., Lie's results would have produced a slope near unity. Perhaps, in the present experimental situation, the knowledge of actual location of the screen played a greater role than in Lie's experiment. The reason why this should have been the case is not entirely clear. The present results are in line with the notion that the visual system reaches a compromise between the competing sets of information about the location of the screen.

Experiment II

The apparent distance of a repeating pattern was measured under two viewing conditions in which convergence was the same but disparities were different. All *S*s served in both conditions. The underlying notion was that if *S*s utilize the information that the repeating pattern was the same in both conditions, the change in disparities could produce a difference in apparent distances. Because disparities produced by a given object are correlated with egocentric distance, a change in disparities can be a cue to distance, if the size of the object is known or assumed to be equal at different distances. However, if the size of the object is not known, disparities produced by an object cannot be a cue to egocentric distance. This argument is consistent with Ittelson's (1960) mathematical derivation of the magnitude of disparity change associated with a change in convergence in a wallpaper situation; i.e., in computing the disparities for different convergence levels the separation of elements in the repeating pattern was assumed to be known. The present argument is that disparities per se cannot provide information for egocentric distance, but disparities and known size together may do so. The hypothesis of Exp. II was that when *S*s encounter a repeating pattern of unknown unit size for the first time, disparities could not determine the apparent distance, but if the same repeating pattern were then presented with altered disparities, apparent distance would be different. Half of the *S*s viewed a repeating pattern through a pair of prism lenses (base in) first and then through a pair of plain glasses. The other half of the *S*s viewed the screen through the plain glasses first and then through the prism lenses. The prediction was that viewing the repeating pattern through prisms or plain glasses makes no difference on the first test period. However, the group whose sequence was plain glasses, prisms will receive an increase in disparities on the second period relative to the first, and hence should report a decrease in apparent distance on the second test period. The group whose sequence was prisms, plain glasses will receive a decrease in disparities on the second test period relative to the first and hence should report an increase in apparent distance on the second test period.

Method

Apparatus.—Both a training and a test apparatus were used in Exp. II. The training apparatus consisted of a target rod, which could be set at various distances in *S*'s median plane, and a slider mechanism similar to that used in Exp. I. The apparatus was designed to allow *E* to give proprioceptive feedback to *S* as to the correct location of the target, while *S*'s hand was on the slider. A plywood sheet supporting the target rod covered *S*'s hand and prevented visual feedback.

The test apparatus was constructed so that information about the mesh size and the distance of the screen would be minimal prior to the experiment. The schematic drawing of the test apparatus is presented in Fig. 2. The apparatus consisted of a wire mesh screen (mesh size 1.3 cm.^{2}), measuring 40.5 cm. high and 83.0 cm. wide, mounted on one side of a rectangular plywood box. At the side of the box, opposite to the screen, a pair of detachable goggles was mounted such that the distance from the front surface of the goggles to the screen was 43 cm. The detachable goggles permitted the insertion of a pair of lenses. Either two plain glasses or two wedge prism lenses (3 diopters, base in) were inserted, depending on the experimental condition. The center of the lenses could be separated to correspond to the interocular distance of *S*. The interior of the box, including the screen, was painted

mat black. (There was no theoretical or methodological reason for having the screen painted black. Because of the difficulty of removing stray paint, the screen was painted black to make the repeating pattern uniform.) Illumination was by means of two 60-w. bulbs located at the two interior corners opposite the screen. The two bulbs gave approximately even illumination and made the screen clearly visible. The screen was viewed against a black wall. As in Exp. I, a small white marker was attached to the screen. Also as in Exp. I, it was possible to insert a thin rod vertically down into S's field of vision to indicate to S the appropriate vergence. The entire apparatus, with the exception of the goggles, was shrouded in black cloth. A guillotine door in front of the goggles controlled the time and duration of viewing of the interior of the apparatus. Thus, no information concerning the mesh size of the screen or the true distance of the screen was available to S prior to the experiment.

Experimental design.—The sequences of No Prism and Prism viewing conditions formed the major experimental variable. The two possible sequences were the basis for two independent experimental groups. One group (No Prism First) viewed the screen first through plain glasses and then through the prism lenses. Another group (Prism First) viewed the screen first through the prisms, then through plain glasses. The Ss were randomly assigned to one of the two groups.

Procedure.—The interocular distance of each S was measured and recorded. The S was seated at the training apparatus and instructed to grasp the handle of the slider mechanism and to place his chin in the chin rest. The S was asked to look at the tip of the vertical rod and set the handle of the slider beneath it. Following each trial, E provided feedback by moving the slider to the correct location beneath the rod, and S returned the slider to the starting position which was approximately below the eyes. Rod positions were 20, 30, 40, and 50 cm. from the bridge of the nose of S. Rod positions were presented randomly in five blocks to constitute 20 trials. The training period lasted 4–5 min. This training procedure was used because an otherwise identical experiment gave ambiguous results.

Following adjustment of the goggles for interocular distance and insertion of the appropriate lenses, S was seated at the test apparatus. The S was instructed to keep his head steady against the goggles and to describe what he saw on the screen when the guillotine door was raised. This procedure was continued until either S met the necessary criteria for each condition or S was rejected from the experiment. With perfect linkage between accommodation and convergence, S should see a single marker on the screen in the No Prism condition and two markers two grid units apart in the Prism condition. When S perceived the marker(s) in this manner, the vergence was the same for both conditions. However, not all Ss fulfilled this requirement. Hence, the following procedure was used to select Ss.

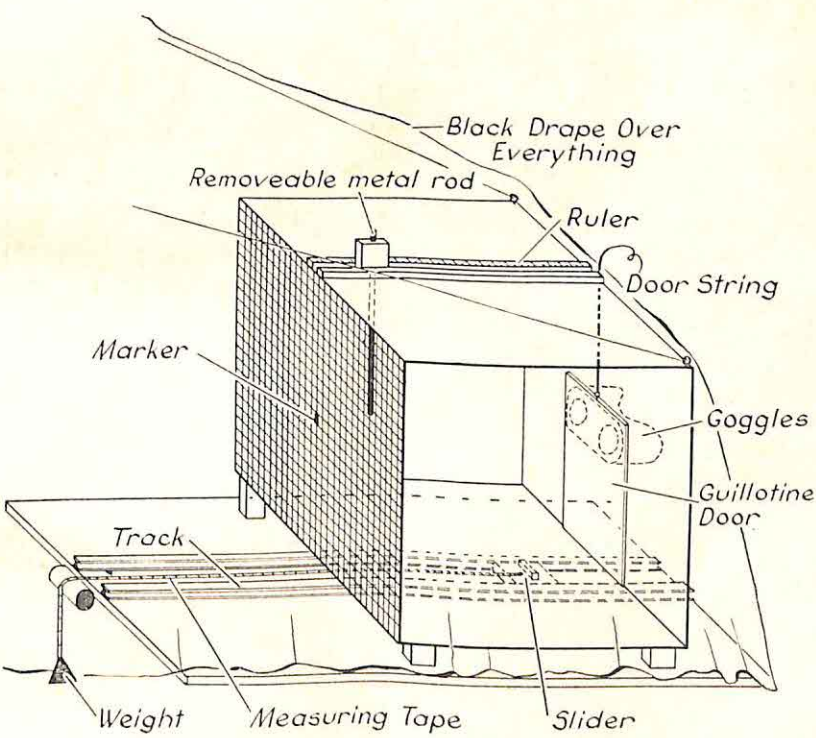

FIG. 2. Schematic drawing of the test apparatus used in Exp. II.

In the No Prism condition, *S*s were required to give three consecutive reports of a single marker in a maximum of 10 exposures of the screen. In the Prism condition, *S*s were required to meet one of two criteria. The first criterion was satisfied if he gave two consecutive reports of two markers two grid units apart in three exposures. If *S* failed to meet the first criterion, an attempt was made to train him to use the required convergence before he developed a tendency to use a different convergence level. To accomplish this, three steps were involved: (*a*) A rod was introduced at the appropriate distance computed from the interocular distance of *S* and the prism characteristics. The guillotine door was raised, and *S* was asked to look at the tip of the rod and to describe what he saw on the screen. (*b*) The door was opened, and if *S* reported the required separation, the rod was removed. (*c*) The guillotine door was opened without the rod. If *S* reported the required separation in Step *a* in two consecutive exposures out of five, he proceeded to Step *b*. In Step *b*, if *S* maintained the required separation after removal of the rod, he proceeded to Step *c*. If not, Step *b* was repeated a maximum of three more times. If *S* saw the required separation in Step *c*, he went to the test period. If not, he went back to Step *b*. If he did not report the required separation in the second *b*–*c* sequence, he was excluded from the experiment. The *S*'s vergence was checked further by asking him, when he saw two markers, to close one eye and report whether the seen marker was on the right or left of the unseen marker. This was done to insure that *S* was not perceiving two markers by converging beyond the screen.

If the criteria for the Prism or No Prism condition were satisfied, *S* proceeded with the test period, and he was asked to set the handle of the slider at the apparent distance of the screen. Settings were made only when *S* saw two markers two grid units apart in the Prism condition, or a single marker in the No Prism condition. Following each setting, *S* returned the slider to the starting position. The intervals between the training and test periods, and between the two experimental conditions, were each approximately 1 min. in length.

Subjects.—A total of 62 *S*s participated in the experiment. Of these, only 31 *S*s met the necessary criteria for both experimental conditions. The criteria appeared more difficult to attain for those *S*s who were in the Prism First group than for those in the No Prism First group. Only 15 out of 39 *S*s met the criteria in the Prism First group, whereas 16 out of 23 *S*s met the criteria in the No Prism First group. One *S* in the No Prism First group was replaced because of the large deviation of his apparent-distance setting from the group mean (3.5 and 4.6 standard deviations above the mean of the remainder of the group for the No Prism and Prism settings, respectively). Of the 30 *S*s used in the analysis, there were six men and nine women per group. All *S*s were university undergraduates.

Results and Discussion

The basic data for analysis were the means of five measurements of apparent distance obtained for each *S* in each viewing condition. The means and standard deviations of apparent distance for each condition and each group (Prism First and No Prism First) are presented in Table 2. A two-way (2×2) analysis of variance, with correlated measures on one factor was performed on the scores. The analysis found a significant interaction ($p < .001$) of Groups × Test Periods, F (1, 28) $= 14.94$. The main effects of Groups and the Test Periods were not statistically significant. Examination of the table reveals the source of interaction. The difference between Prism and No Prism conditions was negligible on the first viewing, as can be seen from the approximately equal means for the first viewing in both groups. Differences between the groups emerged in the second viewing. Viewing through prisms resulted in a smaller apparent distance than viewing through plain glasses.

The results support the hypothesis that when *S*s encounter a repeating pattern of unknown unit size for the first time, disparities do not determine the apparent distance. However, when the same repeating pattern is presented with different disparities, the apparent distance shifts accordingly. The results partially confirm Ames' original observation (Ittelson, 1960), but the generalization about the relationship between the apparent distance and disparities should be limited to a situation in which *S*s are familiar with the repeating

TABLE 2

MEANS AND STANDARD DEVIATIONS OF APPARENT DISTANCES OF FIRST AND SECOND VIEWING FOR NO PRISM FIRST AND PRISM FIRST GROUPS IN EXPERIMENT II

Group	Measure	Test period: First viewing	Test period: Second viewing
No prism first	$\bar{X}$	38.24 (no prism)	35.47 (prism)
	SD	4.84	4.81
Prism first	$\bar{X}$	38.41 (prism)	40.62 (no prism)
	SD	4.82	3.73

pattern. The finding is in line with the notion that disparities per se are not a cue for egocentric distance.

The fact that a greater proportion of *S*s was rejected from the Prism First condition than the No Prism First condition may have introduced a sampling bias and suggests a need to qualify the conclusion of Exp. II. Perhaps, the small curvature distortion introduced by the prisms or the diplopic images of the marker made the task difficult. However, the reason why the same task became easier after viewing the screen with the plain glasses is not clear. If an optical device can be made to eliminate the slight curvature distortion, the difference in proportions of rejection from the two groups may become smaller. Whether the mean apparent distances obtained in the two groups would remain the same with such a device cannot be answered from this experiment.

General Discussion and Conclusions

The two experimental settings investigated were those of Ames' demonstrations (Ittelson, 1960), from which he argued that the changes in disparities rather than convergence were responsible for the wallpaper phenomenon. Experiment I required a dissociation of the usual linkage of convergence and accommodation. Experiment II required the converse, namely, a strong linkage between accommodation and convergence. Given these different requirements, a monitoring of vergence was clearly required in both experiments. Hence, a "marker" was used. However, the marker might have introduced a difficulty in meeting the above requirements in that the marker can serve as a fusional stimulus. Whatever the reasons for the difficulties, more than half of the *S*s who participated in Exp. I and about half of the *S*s in Exp. II failed to meet the requirements.

The chief results obtained from *S*s who fulfilled the requirements were somewhat different from those observed by Ames. In Exp. I, unlike Ames' observation, the two viewing conditions in which a uniform repeating pattern was placed on a Vieth-Mueller circle yielded changes in apparent distances as a function of convergence distances. The decrease in the apparent distances as convergence level increased was not as steep as that of the pattern on a frontal parallel plane; but, nevertheless, the decrease was reliable. In Exp. II, when a convergence level was kept constant but disparities differed, apparent distance was a function of disparities only when *S*s developed assumptions about the size of the units forming the repeating pattern.

With regard to the question about the correct explanation for the wallpaper phenomenon, the results of the present study suggest that the explanation does not lie in a single cue. In the two experimental situations, the visual system has several means of determining apparent distance. Hence, the situations are that of cue conflict. In Exp. I, no attempts were made to eliminate the information about the actual location of the screen or the size of the wire mesh. One set of information consisting of the knowledge of the actual distance, the retinal image size and the known mesh size together, and the accommodation cue would bias the perceived distance toward the actual location. Another set of information opposing this bias consists of the changes in the convergence level and the changes in disparities plus the knowledge of wire mesh size together in the Frontal Parallel Plane condition, or the convergence changes alone in the two Vieth-Mueller circle conditions. The fact that the Frontal Parallel Plane condition yielded slopes of the regression lines for apparent distance on convergence distance closer to unity than slopes from the two Vieth-Mueller Circle conditions is consistent with the cue conflict hypothesis. Also consistent with the hypothesis is the fact that in all three conditions the slopes were closer to unity when a particular screen was viewed for the first time. The knowledge of actual location might have played a greater role in determining apparent distance in the second and third experimental sessions as *S*s become more certain about the actual distance of the screen. In Exp. II, in the first viewing condition, whether *S*s viewed the screen with prisms or no prisms, conflict was minimal. The only cues operating were those of accommodation and convergence. However, in the second viewing condition the increase or decrease in the disparities was in conflict with the fact that convergence, accommodation, and retinal image size remained the same. The results of Exp. II support this hypothesis. The contention of this paper is that perceived distance is an outcome of the visual system processing several sets of information and reaching a compromise when there is conflict between two or more sets.

REFERENCES

GOGEL, W. C. The organization of perceived space. In H. L. Teuber (Ed.), *Handbook on sensory physiology*, 1971, in press.

HELMHOLTZ, H. *Physiological optics*. Vol. 3. (Trans. by J. P. C. Southall) Rochester: Optical Society of America, 1925.

ITTELSON, W. H. *Visual space perception*. New York: Springer, 1960.

LIE, I. Convergence as a cue to perceived size and distance. *Scandinavian Journal of Psychology*, 1965, 6, 109–116.

OGLE, K. N. Spatial localization through binocular vision. In H. Davson (Ed.), *The eye*. Vol. 4. New York: Academic Press, 1962.

ONO, H. Some thoughts on different tasks related to size and distance. In J. C. Baird (Ed.), Human space perception. *Psychonomic Monographs*, 1970, **3**, (13, Whole No. 45).

(February 18, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 11–17

SPATIAL PARAMETERS OF EYE–HAND ADAPTATION TO OPTICAL DISTORTION [1]

JOHN C. HAY [2]
University of Wisconsin-Milwaukee

BARRY LANGDON
University of California, Los Angeles

AND HERBERT L. PICK, JR.
University of Minnesota

This study sought to identify the modifiable parameters of eye–hand coordination in the prism-adaptation situation. The most readily modified parameter was found to be a displacement applying equally to all target positions. A magnification parameter could also be partially modified, in the sense that a wider range of hand movements became identified with a fixed eye-movement range. No nonlinear changes in the eye–hand mapping were found.

How closely does the perceptual adaptation to optical distortion compensate for that distortion? The distortion is a rearrangement of the visual stimulus with respect to other stimulus inputs and also with respect to the motor outputs of the individual. The adaptation can be considered to be a kind of "best-fit" inverse rearrangement of the perceptual-motor system, within the degrees of freedom permitted by the modifiable parameters of that system. Through an analysis of the detailed spatial features of the adaptation and whatever mismatch exists between it and the optical distortion, one may hope to identify those modifiable parameters.

The present study performed such an experimental analysis for a well-known type of rapid-acting perceptual adaptation: the adaptation produced by viewing the hand through a wedge prism (Rock, 1966). The resulting stimulus rearrangement can be represented as a remapping of eye positions on hand positions. Figure 1 illustrates this for the case of a base-right (BR) prism. To simplify the analysis, tests are confined to a one-dimensional, left–right range of eye and hand positions. The eye sees the hand move through Positions A, B, C when in fact it is moving through positions A′, B′, C′. Before and after a few minutes of this experience, eye–hand coordination is tested by presenting visible targets (− 3 through + 3) and having *S* point to them with his visually concealed hand (cf. Held & Gottlieb, 1958).

The prismatic distortion of Fig. 1 has the following spatial characteristics: (*a*) The position of the hand relative to the eye's straight-ahead gaze position is *displaced* by an amount, D (B′ − B in Fig. 1). (*b*) The range of hand positions (A′ to C′) is *magnified* relative to the range of eye positions (A to C) by a factor, M. (*c*) The scale of hand positions is subject to a nonlinear stretch or *taper* (A′B′ < B′C′) by a factor, T. In terms of these three parameters, the prism distortion can be described by the following type of equation:

$$\text{Hand Position} = D + (M)\,(\text{Eye Position}) + (T)\,(\text{Eye Position})^2 \quad [1]$$

This type of quadratic equation gives a description accurate to 1% over the range of positions used in the study.

The spatial features of the perceptual adaptation to prismatic distortion were compared with the above features of the distortion itself, for both BR prisms (Fig. 1) and BL prisms (base-left prisms, which have an effect that is the left–right reverse of Fig. 1).

[1] This study was supported by National Institute of Mental Health Research Grant MH-97588 to the University of Minnesota. The experiments were carried out in the psychology laboratory of Smith College. Carol White assisted in one of the preliminary studies cited. The help and advice provided by Joseph Bauer, Jr., and Richard Held are gratefully acknowledged.

[2] Requests for reprints should be sent to John Hay, Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201.

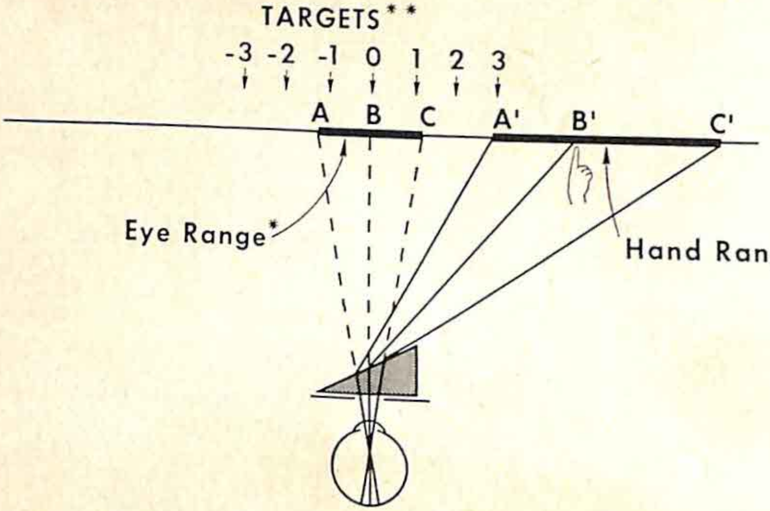

FIG. 1. Geometry of the prism distortion and of the eye–hand coordination test. (One asterisk refers to eye and hand movement ranges during prismatic viewing of the hand. A–C is the range of eye positions; A′–C′ is the range of hand positions. Two asterisks refer to target presented to the eye one at a time during eye–hand coordination test; these targets were not affected by the prism, and the hand was concealed during the response to them.)

In addition, a parallel experiment was done on adaptation to a concave-lens viewing system (Fig. 2), which produced a simpler eye–hand remapping of the following kind:

$$\text{Hand Position} = (M)\ (\text{Eye Position}) \qquad [2]$$

EXPERIMENT I

Method

A modified Held-Gottlieb apparatus was used. The S sat in front of a horizontal table surface, viewing it through a monocular eyepiece with his right eye. His head was positioned by a dental-wax biteboard at a 45° angle so that the center of the

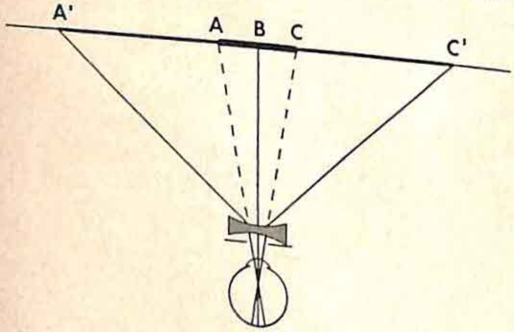

FIG. 2. Geometry of the minifier distortion. (A–C is the range of eye positions; A′–C′ is the range of hand positions.)

table surface was 36 cm. from his eye, measured along a 45° path. This point on the table determined the Position B (or 0) in Fig. 1. The table surface was completely enclosed but could be internally illuminated by two fluorescent lamps. The S could reach his right hand onto the table surface through a curtained slot. A 40-diopter wedge prism, with its base either to S's right or left, could be positioned in the viewing path, producing the distortion illustrated in Fig. 1.

During his prismatic viewing of his hand, S moved his hand back and forth between the left and right visible limits (A and C in Fig. 1). For the eye, these limits were from − 4 to + 4 cm., left and right of the center of the table. For the hand, these limits were from + 10 to + 33 cm., for BR prism, and from − 10 to − 33 cm., for BL prism. These values were empirically verified throughout the experiment. A 10-point plotting of the eye and hand scales gave the following equation for the BR prism:

$$\text{Hand Position} = 16.38 + 2.04\ (\text{Eye Position}) + .1492\ (\text{Eye Position})^2 \qquad [3]$$

The duration of the prism exposure was either 3 min. or 9 min., for two different groups of Ss. A 1-Hz. metronome paced S's hand movements; these were monitored remotely by E using the hand-sensing device to be described below; S was repeatedly reminded throughout the exposure to keep his gaze trained on his right index finger.

Eye–hand coordination test.—The S's ability to place his unseen hand on the visible targets − 3 through + 3 of Fig. 1 was tested in the following way: A 3.8-cm beam-splitter was located between

his eye and the prism; this allowed an undistorted view of seven target lamps physically located above S's head. Their virtual images lay on the table surface, however, at the positions illustrated in Fig. 1. They were uniformly distributed from − 10.6 to + 10.6 cm. relative to the table center; each lamp was .5 cm. in diameter and a dim red when illuminated. When the internal illumination of the table was turned off (and stray light prevented by working in a dark, light-tight room), S could be asked to place his right index finger on the apparent location of an illuminated target lamp, without opportunity to see his hand.

A single test of eye–hand coordination consisted of responding to each of the seven target lamps three times, in three identically ordered blocks. For pretraining, S was run through this test once. He was then given the test again to provide pretest data prior to the prism exposure. Ten seconds after the termination of the prism exposure, terminated by turning off the table illumination, he was given the coordination test a final time, providing posttest data. The target lamp presentations were controlled by an electronic timer and a stepping switch; each lamp was turned on singly for 7.5 sec., with a 2-sec. dark interval between lamps. The order of targets was varied between Ss, in a Latin-square design (see below).

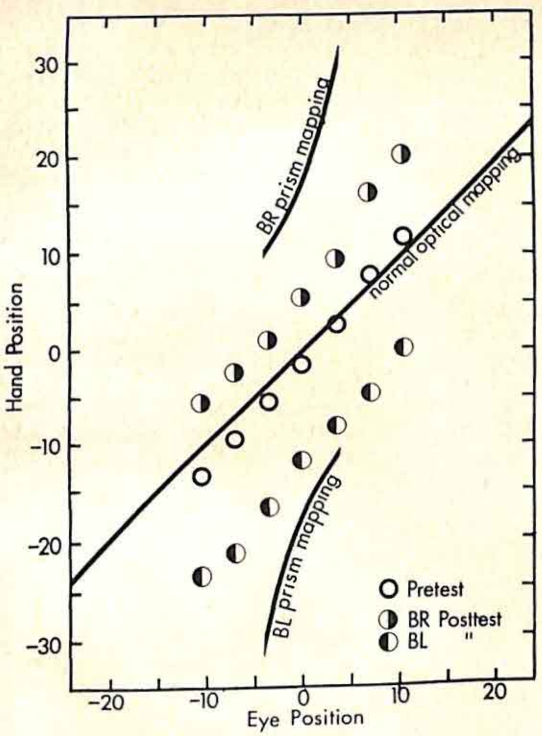

FIG. 3. Effects of 9-min. prism adaptation in Exp. I. (The horizontal and vertical scales (in cm.) measure eye and hand positions, respectively, along the table top; the straight solid diagonal shows the normal optical relationship between the two; the curved solid lines show the BR and BL prismatic remappings of eye and hand; the circles are the empirical eye–hand coordination data, before and after prism exposure.)

The finger placements were registered using the electric-field probe technique of Bauer, Woods, and Held (1969). A 120-cm.-long piece of Teledeltos conductive paper was used, with a lengthwise resistance of 9,500 ohms; a stable parallel-line electric field was established across its length by means of contacts at either end that spanned its 10.2–cm. width, using a Hewlett-Packard (6102A) power supply (.001% regulation, .1-mv. peak-to-peak ripple), such that there was a 100 mv/cm drop over the central 100 cm. of the paper. The S's right index finger held a rubber finger guard with a 1-mm. metal contact; when this probe touched the Teledeltos paper, the field strength at that position could be read out by means of a high input-resistance digital voltmeter (Hewlett-Packard 3430A, 10 megohms input resistance, 1-mv. sensitivity). In practice, this system provided position readings accurate to less than .5 cm.; the accuracy was checked by readings made at 10 positions at the beginning of each experimental session.

Subjects.—Twenty-eight undergraduate women at Smith College served as Ss, 14 with 3 min. of prism exposure, 14 with 9 min. Base-right and base-left prisms were divided equally among them. For each of the four subgroups thereby determined, seven different target orders were used on seven different Ss; the same seven target orders were used in each subgroup, with each target occurring once in each ordinal position.

Results

Figure 3 shows the salient features of the results, as illustrated by the group means for the 9-min. exposure group. An adaptive displacement was found between pretest and posttest, but no variation between targets is apparent to correspond to the magnification and tapering features of the prism distortion. An incidental feature is the greater adaptive displacement found for the base-left prism Ss.

This graphic interpretation was supported by statistical analyses. An analysis of variance (see below) indicated that the only critical factors for adaptation were target position and prism orientation. To analyze the spatial parameters of the adaptations, equations of the form of Equation 1 were fitted individually to each S's pretest and posttest data, using the method of least squares on a PDP-8 computer. These parametric values were

then averaged for all 28 *S*s for the pretest data, and separately for the 14 *S*s in each prism-orientation group. The following three equations present the results. The unit of measurement is the centimeter; the parenthetical values below each equation factor is the standard error for that factor:

Pretest:

$$\begin{aligned}\text{Hand Position} &= \underset{(1.76)}{-2.01} + \underset{(.15)}{1.14}\,(\text{Eye Position}) \\ &\quad + \underset{(.0018)}{.0013}\,(\text{Eye Position})^2\end{aligned} \quad [4]$$

BR Posttest:

$$\begin{aligned}\text{Hand Position} &= \underset{(1.73)}{5.37} + \underset{(.13)}{1.24}\,(\text{Eye Position}) \\ &\quad + \underset{(.0017)}{.0018}\,(\text{Eye Position})^2\end{aligned} \quad [5]$$

BL Posttest:

$$\begin{aligned}\text{Hand Position} &= \underset{(1.87)}{-11.37} + \underset{(.18)}{1.10}\,(\text{Eye Position}) \\ &\quad + \underset{(.0026)}{.0001}\,(\text{Eye Position})^2\end{aligned} \quad [6]$$

The only statistically significant changes are in the displacement factor, *D*. This can best be seen by setting confidence limits for the changes observed in BR and BL groups separately: For displacement, the 95% BR and BL values are $+7.38 \pm 3.73$ and -9.36 ± 4.04, respectively. For magnification, the corresponding values are $+.10 \pm .29$ and $-.04 \pm .40$. For taper, the values are $+.0006 \pm .0037$ and $-.0013 \pm .0056$.

An analysis of variance was performed for the following six factors: prisms (BR vs. BL), test (Pretest vs. Posttest), targets (Positions -3 through $+3$ of Fig. 1), time (3-min. vs. 9-min. exposure), test block (three run throughs of the targets on each test), and *S*s. The results supported the above analyses. The effects relevant to adaptation are those involving the Prism × Test interaction; only two of these were significant. The Prism × Test interaction showed a strong overall adaptation effect, $F(1, 24) = 386.81$, $p < .001$. The Prisms × Test × Target interaction indicated a variation in adaptation for different targets, $F(6, 144) = 2.50$, $p < .05$. The foregoing analysis shows that this was not related in any clear fashion to the features of the prism distortion; the existence of such variation per se has been previously reported (cf. Sekuler & Bauer, 1966). Time of exposure failed to have an effect for the 3-min. and 9-min. values used: the Prism × Test × Time interaction showed $F(1, 24) = .54$. This lack of effect is consonant with the rapid approach to asymptotic adaptation for strong prisms found by Efstathiou (1969). The simple effect of the test factor was significant, $F(1, 24) = 6.14$, $p < .025$. This confirms the greater adaptation for BL prisms indicated by Fig. 3; if the adaptation effects of BR and BL prisms were equal, the difference between pretest and posttests should cancel out. The main effects of targets and prisms were significant, $F(6, 144) = 565.11$, $p < .001$, and $F(1, 24) = 73.30$, $p < .001$, respectively; this merely confirms that hand placements were different for different targets and that adaptation had opposite directions for BR and BL prisms. The only other significant effects were those due to test block interaction with prisms and with targets, $F(2, 48) = 19.55$, $p < .001$, and $F(12, 288) = 2.04$, $p < .05$, respectively.

A further analysis of variance was performed to test for the presence of deviations from linearity in the eye–hand coordination functions. Using the linear trend extraction techniques described in McNemar (1969), the linear trends were found to be highly significant for pretests and posttests of both BR and BL groups; the variance left after extraction of the linear component in no case reached even the .20 level of significance.

Experiment II

It is possible that the strong optical displacement of the prisms in Exp. I somehow masked the magnification and taper effects of the prisms, so far as the detection system that promotes adaptation was concerned.

A form of eye–hand adaptation to magnification has been reported, for judgments of the size of hand-held objects (Rock, 1966). For these reasons, the procedures of Exp. I were repeated using an optical rearrangement that isolated and enhanced the magnification factor (Fig. 2 and Equation 2).

Method

Procedure.—A single concave lens of the kind used to give apartment dwellers a wide-angle view of prospective visitors was used in place of the BL and BR prisms of Exp. I. The optical remapping had the following empirical parameters:

$$\text{Hand Position} = 1.10 + 5.35\ (\text{Visual-Field Position}) \quad [7]$$

(The 1.10 displacement effect was accidental, due to a slight error in centering the lens on the eye's straight-ahead.) The procedures of Exp. I were replicated in detail, except that only one group of 14 *S*s was used, all with 3-min. exposure. The *S*s were drawn from the same population; none had served in an eye–hand adaptation experiment before.

Added observations were made at the end of the experiment in the form of a phenomenological account by *S* of the appearance of her hand during the exposure period, and how she thought this influenced her subsequent behavior.

Results

Figure 4 shows the mean results for all 14 *S*s. A change in the magnification (or "slope") parameter of the eye–hand mapping is visually evident. This is supported by the results of fitting equations of the form of Equation 1 to each *S*'s pretest and posttest data, as in Exp. I:

Pretest:

$$\text{Hand Position} = .01 + 1.20\ (\text{Eye Position}) + .0007\ (\text{Eye Position})^2 \quad [8]$$

Posttest:

$$\text{Hand Position} = -.56 + 1.49\ (\text{Eye Position}) + .0002\ (\text{Eye Position})^2 \quad [9]$$

The only significant change is in the magnification parameter: The 95% confidence limits for change in magnification were

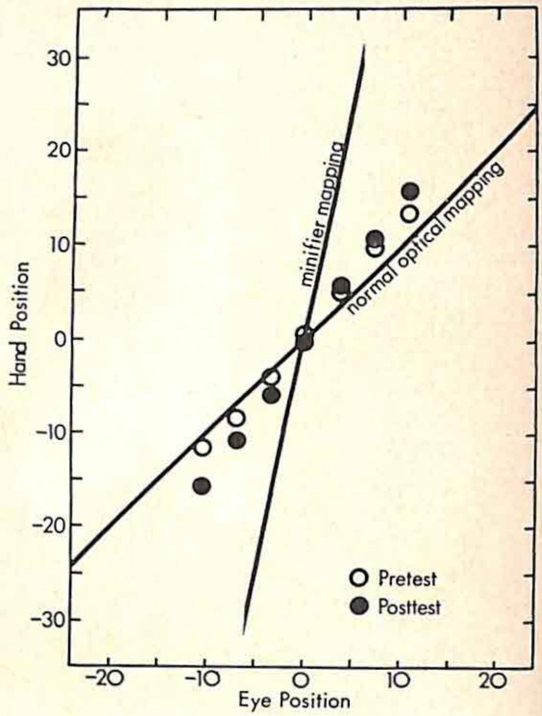

FIG. 4. Effects of 3-min. minifier exposure in Exp. I. (The horizontal and vertical scales (in cm.) measure eye and hand positions, respectively, along the table top; the longer solid diagonal shows the normal optical relationship between the two; the shorter shows the minifier remapping of eye and hand; the circles are the empirical eye–hand coordination data, before and after minifier exposure).

$.29 \pm .09$; those for displacement, $-.57 \pm 1.07$; those for taper, $-.0005 \pm .0015$.

An analysis of variance tested the following factors: test (pretest vs. posttest), target (-3 through $+3$ of Fig. 1), test block (three run-throughs of the targets on each test), and *S*s. The Test × Target interaction showed a differential adaptation for different target positions, reflecting the magnification change, $F\ (6, 72) = 22.19$, $p < .001$. The main effect of the test factor was also significant, $F\ (1,12) = 5.47$, $p < .05$. This reflects the downward pretest-posttest trend in Fig. 4, which was also found in Fig. 3. It represents a tendency for the posttest hand placements to be further to the left than those of the pretests, independent of the optical distortion experienced. The only other significant effect was that due to targets, $F\ (6, 72) = 157.20$, $p < .001$.

A linear trend analysis showed a highly significant linear trend in both pretest and posttest, F (1, 36) = 249.51 and F (1, 36) = 386.41, respectively, both significant at the .001 level. The variance remaining after the extraction of this component was not significant in either case, F (5, 36) = .52 and .39, respectively.

The phenomenological reports of 10 of the 14 *S*s had a striking feature: they claimed that the range of their hand positionings was smaller on the posttest than on the pretest. This verbal report is in exact contradiction to their actual behavior; all 14 *S*s showed an increased posttest range of hand positions, in conformity to the group means of Fig. 4. This indicates that the eye–hand adaptation occurs independently of any understanding that *S* may have of the situation.

Discussion

The most readily modified parameter of eye–hand coordination is an additive constant relating the two. In Exp. I, a mean shift (BR and BL combined) of about 8 cm. was produced in as little as 3 min., corresponding to 12.5° of visual angle at the 36-cm. viewing distance. This adaptation is 51% of the displacement parameter of the prism distortion (Equation 3). In the same experiment, no adaptation was found for the magnification and taper features of the prism distortion.

A change in the magnification parameter of eye–hand coordination can be induced, however. The lens in Exp. II isolated and enhanced the optical magnification experienced by *S*; the result was a .29 adaptive change in the magnification parameter of eye–hand coordination. This amounts to a compensation of about 5% for the magnification of the lens (Equation 7). Such a change may also have occurred in Exp. I; it is not excluded by the confidence limits for that experiment's results.

No nonlinear change in eye–hand coordination was found. It remains possible that if the optical taper of the prism in Exp. I were somehow isolated and enhanced, as was done with magnification in Exp. II, an adaptation to this optical distortion might be induced.

The findings can be algebraically summarized in terms of *adaptation* factors and the following equation:

Change in hand-positioning response =

$$.51(D) + .05(M)\,(\text{Eye Position}) + .00(T)(\text{Eye Position})^2 \quad [10]$$

This equation describes the *change* in hand positioning to a target with a given visual-field position following a 3-min. to 9-min. exposure to an optical distortion with the Parameter Values D (displacement), M (magnification), and T (taper). The numerical values in the equation are the adaptation factors; they were obtained from the observed changes in hand positioning by dividing those changes by the relevant distortion parameter. In deriving the factors, the following assumptions were made: (*a*) if the 95% confidence limits for a factor did not exclude zero, zero was taken as the best estimate of the factor; (*b*) if the 95% confidence limits did exclude zero, the observed mean was taken as the best estimate. The adaptation factors for D and T are based on Exp. I, that for M on Exp. II. Presumably, these factors depend on the duration of exposure to the optical distortion but evidently not over the 3-min. to 9-min. range of Exp. I (no statistically significant changes were found).

Since previous studies of eye-hand adaptation have not provided spatial functions for either the optical distortion used or the adaptations observed, the generality of the above equation cannot be tested by those studies. The .51 adaptive factor for displacement is in general agreement, however, with earlier studies. Efstathiou (1969) discussed the magnitude of this factor and also shows that it rapidly reaches an asymptote, as was found in the present study. Studies in which 100% adaptation is found (cf. Pick & Hay, 1964) have employed much longer, much more diversified exposures to optical distortion.

Does the viewing aperture influence adaptation?—One question neglected so far is whether the limited range of the optical remapping might limit the adaptation. In Fig. 1 and 2, it is shown that the prism and lens had apertures that allowed only the range A–C to be visible during the distorted viewing of the hand (13° of visual angle). The test targets (−3 to +3 in Fig. 1) spanned a wider range (35° of visual angle). It might reasonably be hypothesized that the adaptation would fall off for targets outside the optical distortion range.

The findings on this hypothesis are negative. The linear trends analyses for both experiments showed that a linear function accounted

for the variance of both pretest and posttest data; this would not be the case if adaptation fell off for the two left-most and the two right-most targets, which were outside the visible range during exposure to optical distortion (Fig. 1). Figure 4 underscores this point graphically: the right-most and the left-most targets show greater adaptive changes than do the central targets. Harris (1963) also reported a lack of generalization decrement for targets outside the exposure field.

Replication note.—Experiment I was preceded by two earlier studies that gave similar results. In the first of these, (Hay & Pick, 1967), a limited number of target positions prevented the detailed spatial analysis of the present study. In the second, 16 *S*s were tested under the 3-min. exposure conditions of Exp. I and gave results closely resembling the one presented here.

REFERENCES

BAUER, J. A., JR., WOODS, G. D., & HELD, R. A device for rapid recording of positioning responses in two dimensions. *Behavior Research Methods and Instrumentation*, 1969, **1**, 157–159.

EFSTATHIOU, E. Effects of exposure time and magnitude of prism transform on eye–hand coordination. *Journal of Experimental Psychology*, 1969, **81**, 235–240.

HARRIS, C. S. Adaptation to displaced vision: Visual, motor, or proprioceptive change? *Science*, 1963, **140**, 812–813.

HAY, J. C., & PICK, H. L. Differential eye–hand adaptation to the differential optical displacement of prisms. Paper presented at the Eastern Psychological Association Convention, Boston, April 1967.

HELD, R., & GOTTLIEB, N. A technique for studying adaptation to disarranged hand–eye coordination. *Perceptual and Motor Skills*, 1958, **8**, 83–86.

MCNEMAR, Q. *Psychological statistics.* (4th ed.) New York: Wiley, 1969.

PICK, H. L., JR., & HAY, J. C. Adaptation to prismatic distortion. *Psychonomic science*, 1964, **1**, 199–200.

ROCK, I. *The nature of perceptual adaptation.* New York: Basic Books, 1966.

SEKULER, R. W., & BAUER, J. A., JR. Adaptation to prismatic displacements: Hand position and target location. *Journal of Experimental Psychology* 1966, **72**, 207–212.

(Received February 22, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 18–24

ENCODING TIME FROM ICONIC STORAGE: A SINGLE-LETTER VISUAL DISPLAY [1]

TERRY J. SPENCER [2]

Kent State University

Backward masking functions were obtained for single-letter (SL) visual displays drawn from either 2, 8, or 24 stimulus alternatives. Displays were equated for stimulus energy and equated for percent-correct recognition by varying stimulus energy. All functions asymptoted to their respective No-Mask Controls at a stimulus onset asynchrony of approximately 125 msec. The results suggest that increasing the energy and/or information load of an SL stimulus does not increase the time required for processing from iconic storage.

The recognition accuracy for a very brief tachistoscopic presentation of a single letter (SL) depends on the number of alternative letters possible on any given trial. If *S* has only to choose between 2 letters, he will perform better (with or without correction for guessing) than if he must choose from among 8 or 24 alternatives, assuming constant stimulus energy and overall performance less than 100%. However, the recognition accuracy for the 24-alternative condition can be improved, and equated to that of the 2-alternative, by simply increasing the stimulus energy of the former. This increase in stimulus energy may provide an icon (Neisser, 1967) which is (*a*) of higher quality, i.e., features are both clearer and more abundant, and/or (*b*) more resistant to decay, i.e., features are available for processing over a longer period of time. The present study was designed to determine if the latter alternative is necessary; i.e., does the greater amount of information processed with an increase in stimulus energy require more time for processing from iconic storage.

Relevant to this question are the findings of Swanson and Briggs (1969) and Briggs and Swanson (1970), who obtained an estimate of the rate of information extraction from iconic storage by means of a procedure in which total reaction time is decomposed into components associated with successive processing stages. These investigators conclude that information is sampled from iconic storage at a fixed rate of approximately 1 bit/157 msec. and that only the amount, not the rate, of information sampling is under *Ss'* control.

A more direct measure of the time involved in extracting information from iconic storage may be obtained from backward masking functions for various conditions of SL stimulation. Spencer (1969) and Spencer and Shuntich (1970) have provided evidence that under appropriate conditions, a pattern mask can control the processing or transferring of information from iconic storage to a more permanent form of memory. With this procedure, processing time can be ascertained by determining the range of backward masking, i.e., the time over which a mask impairs performance as compared to a no-mask control. A comparison can be made of the range of backward masking for blocks of SL stimulus presentations differing in the number of stimulus alternatives but either equated for stimulus energy or equated for recognition accuracy. If the improvement in performance resulting from an increase in the SL stimulus energy is at least partially dependent on an icon being available for processing over a longer period of time, the masking range should be correspondingly greater. However, if only the quality and abundance of features available for

[1] The author is grateful to James Wogan for his assistance in data collection and Larry Hawkes for his help in data analysis.

[2] Requests for reprints should be sent to Terry J. Spencer, Department of Psychology, Kent State University, Kent, Ohio 44242.

processing is critical for performance, but not a greater icon persistence, the masking ranges for all conditions should be equivalent.

Method

Subjects.—Four *S*s (two female) were recruited from the graduate and undergraduate population at Kent State University and paid for their services. All had normal or corrected-to-normal vision and were highly practiced in similar tasks prior to beginning the experiment.

Apparatus and stimuli.—All stimuli were presented in a Scientific Prototype Model GB tachistoscope. The front-lighted test stimulus was presented in Field 1 and viewed through a 1.0 neutral-density filter. A small dim back-lighted fixation cross subtending a visual angle of .2° was present in the center of Field 2. The front-lighted pattern mask was presented in Field 3. There were three SL stimulus sets, 2 letters (2–L) consisting of A and X, 8 letters (8–L) consisting of A, H, O, P, T, V, X, and Y, and 24 letters (24–L) consisting of the entire alphabet except for the letters I and Q. Each stimulus group was represented on a separate set of 24 cards, 12 cards for each letter of the 2–L set, 3 for each letter of the 8–L set, and 1 for each letter of the 24–L set. The SL on each card was constructed from black Paratipe (No. 11316) and subtended a visual angle of .2°. Each SL was placed on its card so as to appear slightly above the fixation point, thus eliminating any positional uncertainty. The pattern mask was constructed by placing four Paratipe Ws, back to back and side by side, partially overlapping, to form a pattern similar to three large adjacent Xs. This pattern was placed on the masking card so as to cover the SL test stimulus when both were presented simultaneously. The duration of Field 3 was set at 5 msec. throughout the experiment. The luminances were measured as 24 and 22 mL. for Fields 1 and 3, respectively, and remained constant for the entire experiment. However, the effective luminance of Field 1 was only 2.4 mL. due to the 1.0 density filter.

Procedure.—Backward masking functions were obtained for five conditions of SL presentations; 2–L, 8–L, and 24–L equated for No-Mask Control performance (Equal Performance) along with 8–L and 24–L at the same energy as the 2–L condition (Equal Energy). Energy was defined as duration since the luminance of Field 1 was constant and exposure durations very short. Stimulus onset asynchronies (SOAs) were 50, 75, 100, 125, and 150 msec., and a No-Mask Control, for each SL function. The *S*s were required to identify or guess the stimulus presented on each trial.

Prior to the experimental sessions, the exposure duration was individually determined for each *S* to yield an approximately 85% correct recognition (uncorrected for chance guessing) for each of the three Equal Performance conditions. This equating of performance required an average duration of 4 msec. for the 2–L condition, 8 msec. for the 8–L condition, and 31 msec. for the 24–L condition. Data were collected over 20 half-hour sessions, 15 for the three Equal Performance conditions, and 5 for the 8–L and 24–L Equal Energy conditions. An Equal Energy session occurred after every third Equal Performance session. All Equal Performance sessions contained four 24-trial blocks, one each of the 2–L, 8–L, and 24–L conditions. all at the same SOA, and one No-Mask Control. Conditions were assigned to sessions to achieve the maximum counterbalancing possible within and between *S*s, given the restriction of a single SOA per session. A total, over *S*s, of 288 trials were given for each SL Equal Performance condition at each SOA, with 480 trials being given for each No-Mask Control condition.

Each SL Equal Energy session contained five 24-trial blocks, an 8–L and 24–L occurring twice at two levels of SOA, and one No-Mask Control. The only exception was the third Equal Energy session in which both an 8–L and 24–L No-Mask Control occurred for a total of six blocks. Thus, there were a total, over *S*s, of 192 trials for each SL Equal Energy SOA combination and 288 trials for each of the two No-Mask Controls.

The *S*s received an average of 14 preexperimental sessions to establish the 85% threshold durations and permit practice in all experimental conditions. Feedback designating the correct response was given after each trial during the practice sessions. However, during the experimental sessions feedback of the total number of correct responses was provided only at the end of each block. Prior to each session, *S* was required to dark adapt for 5 min., after which he received 10 2–L No-Mask warm-up trials with feedback. Prior to each block, *S* was told which stimulus set was appropriate and administered 2 practice trials with feedback corresponding to the condition to be presented. The order of trials within a block was randomized by shuffling the stimulus cards. All trials were self-initiated by *S* after *E* gave the ready signal and *S* had the fixation cross in good focus. All responses were recorded by *E*.

Results

The results for the three Equal Performance masking functions and their respective No-Mask Controls are presented in Fig. 1. A repeated-measures analysis of variance (SOA × Stimulus Alternatives × *S*s) indicated significance for SOA, $F\ (5,\ 15) = 41.99$, $p < .01$, stimulus alternatives, $F\ (2,\ 6) = 7.9$, $p < .05$, and the SOA × Stimulus Alternatives interaction, $F\ (10,\ 30) = 4.01$, $p < .01$. Increasing the mask delay resulted in a significant overall increase in recognition accuracy. The significant differences be-

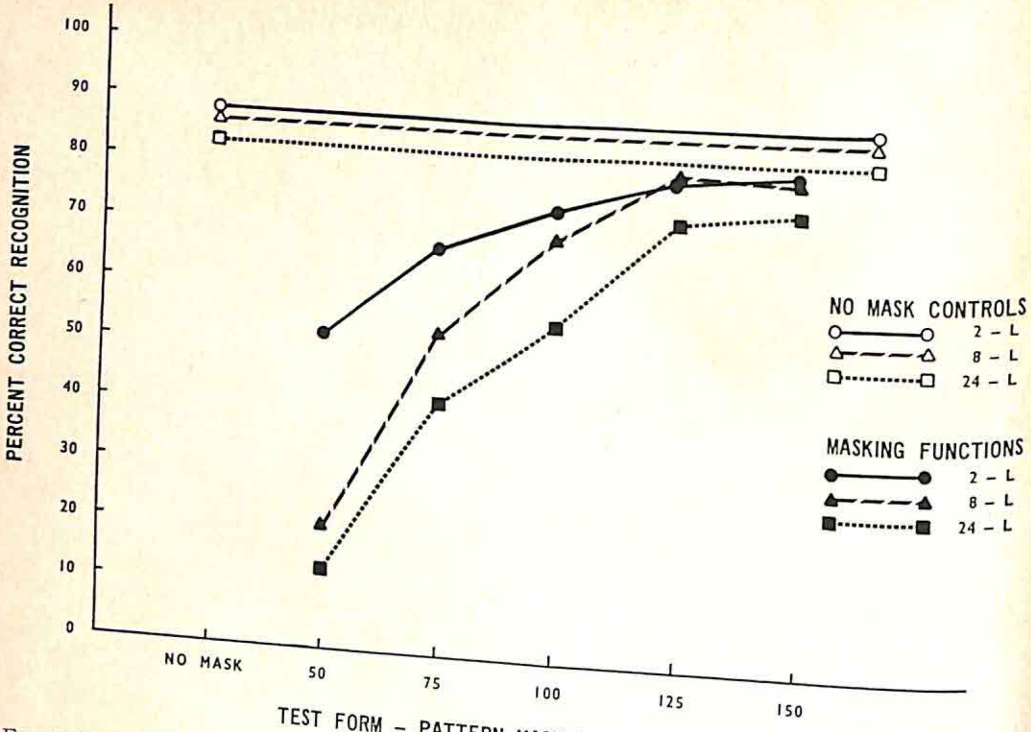

FIG. 1. Percentage of correct identification (uncorrected for chance guessing) as a function of SOA between the test stimulus and pattern mask for the three Equal Performance conditions.

tween stimulus alternatives, as well as the significant interaction, are a result of not correcting the scores for differences in performance that would be expected in the absence of any stimulus information. Thus, essentially chance performance, as at SOA = 50 msec., results in a recognition accuracy of approximately 50% for the 2–L condition, but only 12% for the 8–L, and about 5% for the 24–L condition. Correcting the scores for guessing and repeating the ANOVA resulted in a significant main effect only for SOA, F (5, 15) = 40.12, $p < .01$. A Dunnett's test (Winer, 1962) for comparing all means with a control indicated that all SOAs ≤ 100 msec. were significantly different ($p < .05$) from their appropriate No-Mask Control, while SOAs ≥ 125 msec. were not. This was true for both uncorrected and corrected scores.

The results for the three Equal Energy masking functions are presented in Fig. 2 (the 2–L function is reproduced from Fig. 1). An analysis of variance indicated significance for the main effects of SOA, F (5, 15) = 18.22, and stimulus alternatives, F (2, 6) = 788.93, both $ps < .01$. Both main effects were also significant at $p < .01$ with corrected scores. A Dunnett's test for the 8–L and 24–L Equal Energy functions, for both uncorrected and corrected scores, indicated that each function was significantly different ($p < .05$) from their No-Mask Control at SOA ≤ 75 msec., as compared to SOA ≤ 100 msec. for the 2–L function. This result would suggest that the 8–L and 24–L functions asymptoted faster than the 2–L, but this conclusion may be unwarranted due to the greater variability (partly a result of less data) in the 8–L and 24–L Equal Energy conditions favoring the null hypothesis. In either case, it would seem safe to conclude that all five SL functions asymptoted to their respective No-Mask Controls by SOA = 125 msec. Furthermore, if there are any differences in the masking ranges, they are small (less than 25 msec.) and, most significantly, are not a result of the 8–L and 24–L Equal Performance func-

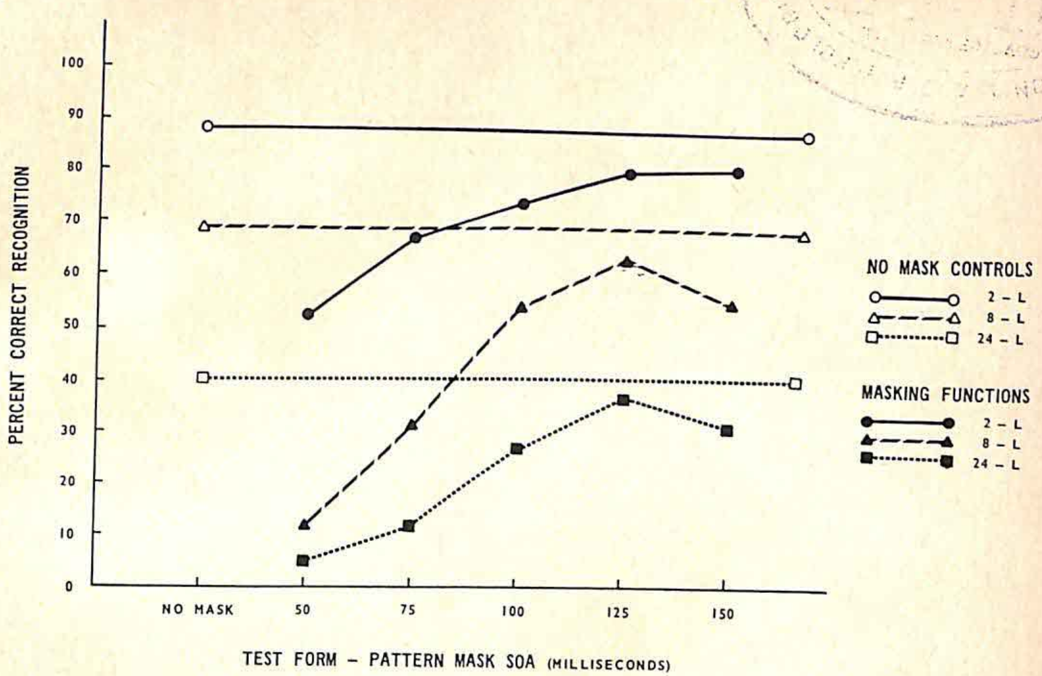

FIG. 2. Percentage of correct identification (uncorrected for change guessing) as a function of SOA between the test stimulus and pattern mask for the three Equal Energy conditions.

tions requiring a greater mask delay than the 2–L function.

An analysis of variance (Energy × SOA × *S*s) for the two 8–L functions considered alone indicated significance for the main effects of energy, $F\ (1, 3) = 11.57$, $p < .05$, and SOA, $F\ (5, 15) = 29.39$, $p < .01$. A similar analysis for the two 24–L functions indicated significance for the main effects of energy, $F\ (1, 3) = 29.79$, $p < .01$, SOA, $F\ (5, 15) = 21.82$, $p < .01$, and the Energy × SOA interaction, $F\ (5, 15) = 4.49$, $p < .01$. As expected, increasing the energy of the stimulus significantly improved the percent-correct performance for both the 8–L and 24–L conditions. The significant Energy × SOA interaction was a result of the pattern mask being equally effective for both 24–L conditions at SOA = 50 msec. and a very different asymptote at larger SOAs. Comparable results were obtained with corrected scores.

A calculation of I_t (Garner & Hake, 1951) for each of the five No-Mask Control conditions indicated .5 bits transmitted for the 2–L condition, 1.7 and 1.6 bits transmitted for the 8–L and 24–L Equal Energy conditions, respectively, and 2.4 and 3.6 bits for the 8–L and 24–L Equal Performance conditions, respectively.[3] Increasing the energy of the stimulus resulted in a significant increase in I_t for both the 8–L, $t\ (3) = 5.03$, $p < .02$, and 24–L, $t\ (3) = 5.27$, $p < .02$, conditions without an increase in the masking range. Thus, for the 24–L Equal Performance condition, over 3 bits of information were processed with the SOA equal to 150 msec. Although the exact value of this I_t may be uncertain due to the effects of information bias (MacRae, 1970), there can

[3] I_t was calculated separately for each *S*. Since *n* was small relative to the number of cells in the 24–L matrix, a correction procedure, similar to that used by Garner (1962, p. 80), was devised. All off-diagonal cells with only one response were assumed random, and the total proportion of random responses within each response category were equally divided among all the off-diagonal cells containing less than two responses. This procedure would yield an I_t varying from close to 0 bits for random guessing to 4.58 bits for perfect responding. The effect of this procedure was to reduce I_t, e.g., from 4.1 to 3.6 bits in the 24–L Equal Performance No-Mask Control condition.

be little doubt that it far exceeds 1 bit due to the large number of perfect responses (75% of the responses were in the 24 diagonal cells) in the 576-cell confusion matrix for the 24–L Equal Performance condition at SOA equals 150 msec.

A conversion of percentage scroes (uncorrected for chance guessing) to a d' measure for the No-Mask Controls by means of Elliott's forced-choice tables in Swets (1964) indicated an average sensitivity of approximately 1.7, 1.9, and 1.6 for the 2–L, 8–L, and 24–L Equal Energy conditions, respectively, and 2.7 and 3.0 for the 8–L and 24–L Equal Performance conditions, respectively. The increase in d' with an increase in stumulus energy was significant for both the 8–L, $t\ (3) = 5.79$, $p < .02$, and 24–L, $t\ (3) = 3.22$, $p < .05$. This pattern of scores is very similar to those of I_t except for the 2–L condition. The d' measure was primarily sensitive to the energy of the stimulus, which is consistent with Signal-Detection theory. The conversion tables used are designed to "control" for the effects of a differential number of response categories on performance.

Discussion

The near equivalence of the masking ranges for all conditions suggests that increasing the energy and/or information load of an SL stimulus does not require additional processing time from iconic storage. This conclusion rests on the premise that backward masking by a pattern mask does indeed control encoding from iconic storage. Evidence for this interpretation has been provided by Spencer (1969) and Spencer and Shuntich (1970).

This result would seem to be inconsistent with the conclusion drawn by Swanson and Briggs (1969) and Briggs and Swanson (1970), that Hick's law of the speed-accuracy trade-off in reaction time experiments is manifested in the amount of information extracted from the iconic storage stage of processing. If the Swanson and Briggs and Briggs and Swanson estimate, that the rate of information extraction is limited to 1 bit per 157 msec., were a general limitation on information sampling from iconic storage, the 8-L Equal Performance condition representing 2.4 bits would have required approximately 375 msec. of uninterupted viewing, and the 24-L (3.6 bits) condition over 500 msec. However, the available information in all conditions appears to have been encoded from iconic storage within 150 msec.

The nature of the encoding process is less clear. The present results would appear to be in agreement with either a template or a parallel feature-testing model of pattern recognition, assuming that full identification of the stimulus occurs within the icon's duration. The phenomenal report by *S*s would seem to favor a parallel feature-testing model. The stimulus input often appeared as a fragmented or distorted cluster of features which *S* learned to decipher during the preexperimental sessions. However, the fact that *S*s required a period of learning could also be interpreted as providing the necessary experience for developing new templates suitable for the "different nature" of low energy stimulus inputs.

It is possible that *S*s did not fully identify the stimulus while it was in iconic storage but transfered it into a more permanent visual storage system, ie., a visual image, which cannot be affected by a subsequent pattern mask. The recognition process, e.g., feature testing, template matching, etc., may have occurred subsequently. Support for the concept of a visual image is provided by the ability of *S*s to report the nature of the visual cues for a stimulus which has been followed by a pattern mask at a short SOA (e.g., 125 msec.). Presenting an A will often result in *S* correctly identifying it, and describing the input as "an upside down V," etc., *S* may see only a fragmentary input but "remember" what it looked like after the mask's arrival. Evidence provided by Kohlers and Katzman (1966), Potter and Levy (1969), and Eriksen and Collins (1969) suggests that successive stimuli presented at rates of 200 to 250 msec. per stimulus can be fully processed during the duration of presentation. However, all three studies found performance decrements at input rates of 125 msec. per stimulus. It could be argued that 125 msec. is insufficient time to fully process even a simple visual display, and performance will suffer if additional processing requirements are placed on a short-term visual storage, as would occur with successive stimuli. This would not be a limitation in the present study and any additional processing time required for identification may have been provided by transferring the icon into a short-term visual storage. In either case, the performance in-

crease with a higher energy stimulus was not a result of a longer lasting icon, but rather an icon of higher quality.

It is important to point out that the estimates of processing time from iconic storage in the present study are relatively independent of the choice of response measure. Analyses made with or without correction for guessing, or by converting percentages to d' scores by means of Elliott's forced-choice tables in Swets (1964), or by calculating the information transmitted (I_t), lead to essentially the same conclusions concerning the masking range for each function. The differential effect these measures have on the shape of the masking functions are small compared to the large overall effect of SOA. However, for other comparisons, such as performance across No-Mask Control conditions, the choice of performance measure can be critical.

A comparison of response measures for the No-Mask Controls reveals that equal performance with respect to uncorrected percentage scores (as shown in Fig. 1) did not result in equal I_t or d' measures. This was not surprising since percentage scores only reflect the relative performance within a given stimulus alternative condition without regard to the difference in difficulty between that condition and some other condition. Applying a two-state correction for guessing, $P(c) = P(C) - [1 - P(C)]/(N - 1)$ where N equals number of response alternatives, $P(C)$ = obtained percent correct, and $P(c)$ = "true" percent correct, reduces but does not eliminate this discrepency between percentage scores and the I_t and d' measures. The two-state model corrects for guessing on those trials representing a complete failure in recognition but does not provide for the probability of a "true" recognition depending upon the number of stimulus and/or response alternatives, or recognition occurring with different degrees of confidence. The d' measure, on the contrary, attempts to assess performance on a more "absolute" scale of signal detection, independent of the number of alternatives or *S*s' decision criterion. I_t also attempts to do this but is more restrictive with respect to the experimental conditions which permit an adequate measure of channel capacity.

The near equivalence of the d' scores for the three Equal Energy conditions tends to support the validity of the Theory of Signal Detection (TSD) as an appropriate model for comparing performances across conditions with different numbers of stimulus and response categories (when performance $< 100\%$). These results are in agreement with the description and analysis of the Miller, Heise, and Lichten (1951) data by Green and Birdsall (1964), who showed that essentially the same d' was obtained for a wide range of vocabulary sizes for a given speech-to-noise ratio. The increase in d' with an increase in stimulus energy, as in the 8-L and 24-L High Energy conditions in the present study, is also in accord with TSD theory.

Garner (1962) has also been able to fit I_t measures to the data of Miller, Heise, and Lichten (1951). The pattern of I_t results for the present study were similar to those of d' except for the 2-L condition. It seems likely that this was a result of the stimulus and response uncertainty in the 2-L condition being too low to permit an adequate measure of "channel capacity." The fact that d' in the 2-L condition was not restricted in this way, as well as the much greater ease of computing d' from forced-choice data, would seem to favor its use over I_t as a response measure. However, the present study was not designed to specifically test these models or their assumptions and any conclusions concerning the relative appropriateness of each measure should be viewed with caution.

REFERENCES

Briggs, G. E., & Swanson, J. M. Encoding, decoding, and central functions in human information processing. *Journal of Experimental Psychology*, 1970, **86**, 296–308.

Eriksen, C. W., & Collins, J. F. Visual perceptual rate under two conditions of search. *Journal of Experimental Psychology*, 1969, **80**, 489–492.

Garner, W. R. *Uncertainty and structure as psychological concepts*. New York: Wiley, 1962.

Garner, W. R., & Hake, H. W. The amount of information in absolute judgments. *Psychological Review*, 1951, **58**, 446–459.

Green, D. M., & Birdsall, T. G. The effect of vocabulary size on articulation score. In J. Swets (Ed.), *Signal detection and recognition by human observers: Contemporary readings*. New York: Wiley, 1964.

Kohlers, P. A., & Katzman, M. T. Naming sequentially presented letters and words. *Language and Speech*, 1966, **9**, 84–95.

MacRae, A. W. Channel capacity in absolute judgment tasks: An artifact of information bias? *Psychological Bulletin*, 1970, **73**, 112–121.

Miller, G. A., Heise, G. A., & Lichten, W. The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, 1951, **41**, 329–335.

NEISSER, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.
POTTER, M. C., & LEVY, E. I. Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, 1969, **81**, 10–15.
SPENCER, T. J. Some effects of different masking stimuli on iconic storage. *Journal of Experimental Psychology*, 1969, **81**, 132–140.
SPENCER, T. J., & SHUNTICH, R. Evidence for an interruption theory of backward masking. *Journal of Experimental Psychology*, 1970, **85**, 198–203.
SWANSON, J. M., & BRIGGS, G. E. Information processing as a function of speed versus accuracy. *Journal of Experimental Psychology*, 1969, **81**, 223–229.
SWETS, J. A. *Signal detection and recognition by human observers: Contemporary readings*. (Ed.). New York: Wiley, 1964.
WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received February 24, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 25–29

PERCEPTION OF ROTATION IN FIGURES WITH RECTANGULAR AND TRAPEZOIDAL FEATURES [1]

MYRON L. BRAUNSTEIN [2]

University of California, Irvine

The figural bases for the effect of perspective changes on accuracy of direction of rotation judgments was investigated. Forms varying in angles of contour convergence, relative extent of vertical contours, and horizontal position of the axis of rotation were displayed in rotation about a vertical axis at five perspective levels. Accuracy, averaged across perspective levels, was ordered primarily by angle relationships. Accuracy was greatest for forms containing right angles, even when the vertical contours were unequal, and lowest for forms in which one vertical contour was enclosed in acute angles and the other in obtuse angles, even when the vertical contours were equal. Within angle relationships, accuracy was generally greater when the vertical contours were equal.

Accuracy of judgments of direction of rotation for figures rotating about an axis perpendicular to the line of sight has recently been studied by Power (1967), Braunstein and Payne (1968), and Murch (1970). Braunstein and Payne employed computer-generated motion picture sequences representing projections onto a plane of figures rotating about a vertical axis. Variations in the distance between the projection point and the axis of rotation were used to simulate variations in viewing distance. In polar projections of rotating figures, the distance of the projection point from the axis of rotation affects both the horizontal and vertical proportions of the projected figure. In the Braunstein and Payne study, some of the stimuli were generated with special projections in which only the horizontal or only the vertical dimensions were affected by the distance of the projection point. Accuracy of direction judgments increased with perspective, defined as $(E + w)/(E - w)$, where E is the distance from the projection point to the axis of rotation and w is one half the width of the original figure. Changes in the vertical dimensions of the projected figures during rotation were found to be more important than changes in the horizontal dimensions in determining the accuracy of direction judgments. When the perspective ratio was greater than 1.0 and less than 1.75 and distance information from sources other than changes in the shape of the image projected on the retina was minimized, accuracy was greater for rectangles than for trapezoids, circles, or ellipses. In polar projections of rotating rectangles or in special projections in which only the vertical dimensions of the projected figure are varied, right angles are projected as acute angles when the enclosed side moves closer to *S* and as obtuse angles when it moves further away. The hypothesis was advanced that *S*s are particularly sensitive to deviations from right angles in a rotating figure.

The present study attempts to define more precisely the information used by *S*s in judging direction of rotation both by altering the availability of potential sources of information and by rendering information from these sources misleading. The two sources of information of primary concern are the angles which the vertical contours form with the horizontal contours and the relative extents of the vertical contours. These sources are completely correlated when rectangles and trapezoids are studied. A series of forms was therefore developed for the present study which

[1] This research was supported by National Science Foundation Grants GB 5545 and GB 8251. The author is indebted to P. Michael Kemp for his assistance in the design of the stimuli and to John W. Payne for his assistance in the conduct of the experiment and the preparation of this manuscript.

[2] Requests for reprints should be sent to Myron L. Braunstein, School of Social Sciences, University of California, Irvine, California 92664.

combined geometrical characteristics of rectangles and trapezoids. These characteristics were encompassed in three factors: (*a*) angles between contours; (*b*) extents of vertical contours; and (*c*) position of the axis of rotation with respect to the vertical contours. The axis of rotation was either located at the center of the form or was displaced horizontally. This displacement can be regarded as a characteristic of a trapezoid when the trapezoid represents a perspective view of a slanted rectangle. In such a view, the projected center of the rectangle is closer to the shorter vertical side.

Method

Subjects.—The *S*s were 48 students in an introductory psychology class who participated as part of a course requirement. Vision of 20/40 or better in the preferred eye was required, on the basis of a Snellen eye chart test. Twenty-four *S*s were randomly assigned to each of the two conditions of symmetry described below.

Stimuli.—The stimuli were 180-frame, 16-mm., computer-generated motion picture sequences representing line patterns (forms) rotating 360° about a vertical axis. The forms combined parts of a rectangle and a trapezoid of equal width. There were 36 forms constructed on the basis of three factors: angles (both right, acute and obtuse, obtuse and acute, both acute, both obtuse, or no angles), sides (equal, right side shorter, or left side shorter), and symmetry versus asymmetry of the horizontal bar about the axis of rotation. The 18 forms generated from the first two factors are shown in Fig. 1. The axis of rotation was at the midpoint of the horizontal bar in the symmetrical condition. In the asymmetrical condition, one third of the horizontal bar was to the right of the axis of rotation. The sequences were generated using each of five mathematical projection points (see Braunstein, 1966), located at distances from the axis of rotation equal to 2.5, 5, 10, 20, and 40 times the width of the form. The corresponding perspective ratios were 1.5, 1.22, 1.11, 1.051, and 1.025.

Apparatus.—A 16-mm. motion picture projector (L-W Model 224-A), operated at 24 frames/sec, was used to display the stimulus films. Two of the shutter blades were covered to avoid the perception of multiple images noted by Braunstein and Coleman (1966). The film was projected onto a translucent (Polacoat) screen. The *S* viewed the screen monocularly from the opposite side at a distance of 1.8 m. through a tube restricting his field of view to a circle 1.2 m. in diameter. All of the patterns were .72 m. wide when shown in the frontal parallel plane. The *S*'s response device consisted of two microswitches mounted in a table under the viewing

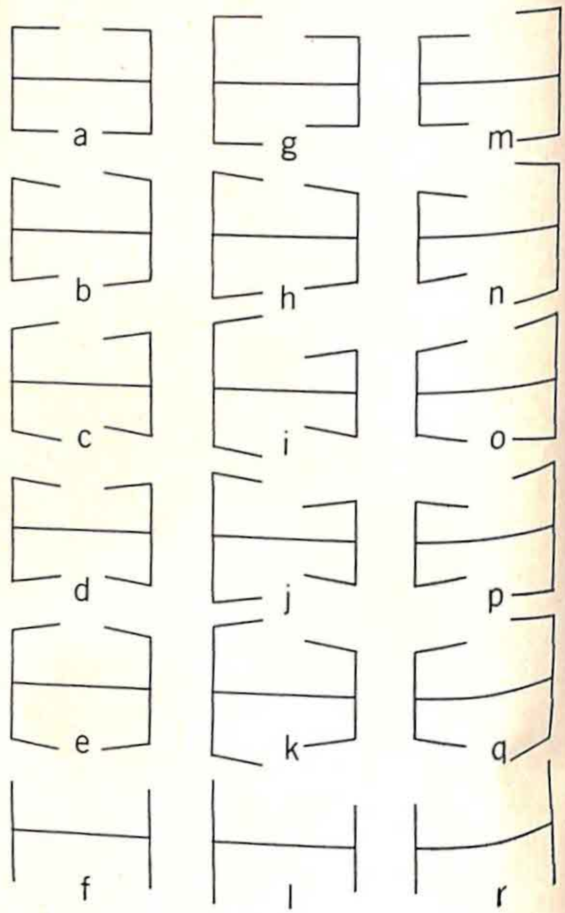

FIG. 1. The 18 forms. (The *S*s viewed white forms against a black background. The axis of rotation was either midway between the vertical lines or displaced to the right.)

tube, 16.5 cm. apart. An 18.4 cm. × 13.8 cm. white metal rectangle with a horizontal crossbar, mounted on a reversible 8-rpm motor, was located at eye level to the right of the viewing tube.

Procedure.—The 90 symmetrical sequences were arranged in random order. The random order was divided into four blocks. Twelve arrangements of these blocks were presented to 12 *S*s. The other 12 *S*s were shown the same arrangements of the sequences with the projector operating in reverse. This procedure yielded a different order of presentation of the stimuli for each *S*, with stimulus position counterbalanced across *S*s. The same procedure, beginning with a new random order, was used to generate 24 orders of presentation for the 90 asymmetrical sequences.

The two directions of rotation, as well as oscillation, were demonstrated with the metal rectangle. The *S* was instructed to hold down the button on the right when the plane appeared to be rotating clockwise, looking from the top. He was asked to change buttons immediately when the direction appeared to change, but to have one of the buttons

depressed at all times (see Braunstein & Payne, 1968). The *S*s practiced the response procedure with the metal rectangle serving as the stimulus. The first five sequences were treated as practice sequences and were repeated later in the experiment. The stimulus sequences lasted 7.5 sec. and were separated by 5-sec. periods.

RESULTS

Each *S*'s response to a stimulus sequence was placed in one of three categories: accurate, inaccurate, or oscillation. A response was considered accurate if *S* held down the button corresponding to the displayed direction of rotation for the entire sequence, after the initial .5 sec. of his response. A response was considered inaccurate if *S* held down the inappropriate button for the entire sequence, after the first .5 sec. of his response. A response was considered an oscillation if *S* pressed both of the buttons during the stimulus sequence, after the first .5 sec. of his response. (The decision to exclude the first .5 sec. of each response, based on previous work with this response procedure, affected the classification of 1.1% of the responses.)

Accuracy increased with perspective level for all of the forms studied. The proportions of accurate responses for the 18 forms and 2 symmetry conditions, taken across *S*s, were .20, .29, .46, .70, and .92 for perspective ratios of 1.025, 1.05, 1.11, 1.22, and 1.50, respectively. The proportions of oscillating responses were .73, .65, .49, .27, and .06 for the same perspective levels. Table 1 shows the mean proportions of accurate and oscillating responses for the 18 forms in the 2 conditions of symmetry. The dependent variable for the analyses of variance was the proportion of accurate responses made by each *S* to presentations of a form at five perspective levels.

TABLE 1

PROPORTIONS OF ACCURATE AND OSCILLATING RESPONSES

Form (see Fig. 1)	Symmetry about axis of rotation			
	Symmetrical		Asymmetrical	
	Accurate responses	Oscillating responses	Accurate responses	Oscillating responses
a	.79	.17	.78	.17
b	.42	.53	.40	.55
c	.47	.51	.42	.55
d	.69	.29	.67	.30
e	.53	.39	.61	.31
f	.59	.29	.53	.32
g	.63	.34	.65	.32
h	.33	.66	.23	.77
i	.59	.39	.58	.37
j	.50	.47	.47	.48
k	.46	.48	.42	.52
l	.42	.47	.32	.57
m	.67	.30	.70	.27
n	.51	.45	.48	.47
o	.29	.71	.39	.61
p	.57	.37	.53	.44
q	.42	.52	.49	.45
r	.45	.48	.38	.51

Note.—Based on 120 observations (24 *S*s at 5 perspective levels). Remaining responses were inaccurate.

The implications of variations in the angles between contours and in the extents of the vertical contours are not exactly the same for the two conditions of symmetry. In the symmetrical condition, the stimuli in the center column of Fig. 1 differ from those on the right only in the initial orientation of the form with respect to *S*. In the asymmetrical condition, the axis of rotation is closer to the shorter side for the forms in the center column and closer to the longer side for the forms on the right. Analogous statements could be made about angle relationships for the forms in the second and third rows. Separate within-*S*s analyses were therefore conducted for the two conditions of symmetry. A three-way analysis of variance for the symmetrical condition (angles, sides, and *S*s), using an arc-sin transformation of the proportions, showed significant ($p < .05$) main effects for angles, F (5, 115) = 18.71, sides F (2, 46) = 9.62, and the interaction of angles with sides, F (10, 230) = 5.69. A similar analysis for the asymmetrical condition also showed significant effects ($p < .05$) for angles, F (5, 115) = 29.12, sides, F (2, 46) = 17.19, and the interaction, F (10, 230) = 4.95.

DISCUSSION

Accuracy of direction judgments was highest for the form most similar to a rectangle (*a*) and lowest for the forms most similar to a trapezoid (*h*, *o*). (Form designations refer to

Fig. 1.) Intermediate results were obtained for the remaining forms combining features of rectangles and trapezoids. The extremity of the difference in accuracy between rectangles and trapezoids suggests that this difference is based on characteristics of both forms. A detailed examination of the ordinal relationships among the proportions of accurate judgments for the 18 forms provides a clear and consistent basis for the discussion of these characteristics. Direction judgments for forms with right angles were more accurate than those for forms containing any other angles or combinations of angles, including no angles. This relationship held among forms with unequal sides (*g* vs. *h, i, j, k, l*; *m* vs. *n, o, p, q, r*), as well as among forms with equal sides (*a* vs. *b, c, d, e, f*). Probably the most important heuristic used in relative distance perception for line figures is that a change from right angles to acute angles implies that the enclosed side is moving toward *S*, while a change from right to obtuse angles implies that the side is moving away.

The omission of right angles decreased accuracy but probably did not mislead *S* in any particular way. The addition of trapezoidal features, on the other hand, provided *S* with misleading distance information. The most important factor is the enclosure of one side in obtuse angles and the other in acute angles. This reduced accuracy even when the sides were of equal length (*b, c* vs. *a, d, e, f*). The side enclosed by the acute angle tended to appear closer throughout the rotation, at the lower perspective levels. The finding that trapezoids appear to oscillate in a consistent fashion (Cook, Mefferd, & Wieland, 1967) appears to be principally due to the tendency to respond to converging lines as indicating a receding surface, even when more accurate distance or slant information is potentially available from other sources (Braunstein & Payne, 1969). In a complete trapezoid, the presence of acute angles on one side and obtuse angles on the other is a necessary consequence of the inequality of the sides. When these factors were separated, the length of the sides proved less important than the angle relationships. Accuracy was greater for forms having unequal sides and right angles (*g* and *m*) than for forms having equal sides, but with acute angles on one side and obtuse angles on the other (*b* and *c*).

A difference in the length of the two vertical sides, without a difference in the angles associated with the two sides, was sufficient to reduce accuracy of direction judgments. Accuracy was greater for the form with equal sides than for the two corresponding forms with unequal sides, in all four equal angle cases: both right (*a* vs. *g, m*); both acute (*d* vs. *j, p*); both obtuse (*e* vs. *k, q*); and no angles (*f* vs. *l, r*). The interaction of angle differences and side differences is especially interesting. When the sides were unequal and the shorter side was associated with the obtuse angles (*h, o*) as in a complete trapezoid, accuracy was lowest. When the longer side was associated with obtuse angles (*i, n*), accuracy was higher than when the sides were equal and the angles were different (*b, c*), or when the sides were unequal and the angles were omitted (*l, r*). The misleading effects of the angle difference and the misleading effects of the side difference in Forms *i* and *n* seemed to work in opposite directions, resulting in greater accuracy.

The effect of asymmetry of the form about the axis of rotation was small. The ordinal relationships discussed held for all pairs of forms listed above, for both the symmetrical and the asymmetrical displays. When the side which was shorter and associated with obtuse angles was also closer to the axis of rotation, the tendency for *S*s to report oscillation increased.

The present results are consistent with those of Murch (1970), who studied perception of rotation for eight objects rotated about a vertical axis through the object's center. Murch's Objects 2 and 7 are essentially the same as Forms *f* and *l*. His Objects 5 and 6 are similar to Forms *a* and *h*. The mean reversal frequencies found by Murch were 6.6, 11, 17, and 22 for Objects 5, 2, 7, and 6, respectively. The perspective ratio in his experiment was 1.09, computed from the reported object width and viewing distance. At a perspective ratio of 1.1 in the present study, the mean proportions of oscillating responses for Forms *a, f, l* and *h*, in the symmetrical condition, were .08, .17, .62 and .92, respectively. The similar ordering of results for these four types of stimuli is a promising indication of progress in the direction of understanding perception of rotation for plane figures with visible contours. Perspective variations also determine accuracy of rotation judgments when contours are not displayed and when a three-dimensional pattern is represented (Braunstein, 1966). The specific findings of the present investigation cannot be directly generalized to stimuli of that type, and further research should be directed toward determining the sources of information used by *S*s in the absence of well-defined contours.

REFERENCES

Braunstein, M. L. Sensitivity of the observer to transformations of the visual field. *Journal of Experimental Psychology*, 1966, **72**, 683–687.

Braunstein, M. L., & Coleman, O. F. Perception of temporal patterns as spatial patterns during apparent movement. *Proceedings of the 74th Annual Convention of the American Psychological Association*, 1966, **1**, 69–70.

Braunstein, M. L., & Payne, J. W. Perspective and the rotating trapezoid. *Journal of the Optical Society of America*, 1968, **58**, 399–403.

Braunstein, M. L., & Payne, J. W. Perspective and form ratio as determinants of relative slant judgments. *Journal of Experimental Psychology*, 1969, **81**, 584–590.

Cook, T. H., Mefferd, R. B., & Wieland, B. A. Apparent reversals of orientation (perspective reversals) in depth as determinants of apparent reversals of rotary motion. *Perceptual & Motor Skills*, 1967, **24**, 691–702.

Murch, G. M. The perception of rotary motion. *Journal of Experimental Psychology*, 1970, **86**, 83–85.

Power, R. P. Stimulus properties which reduce apparent reversal of rotating rectangular shapes. *Journal of Experimental Psychology*, 1967, **73**, 595–599.

(Received February 24, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 30–33

EFFECTS OF RESPONSE LABELS IN CONCEPT ATTAINMENT [1]

RICHARD L. GOTTWALD [2]

Yale University

A comparison of concept attainment difficulty was made between two conditions: neutral responses (alpha vs. beta) and positive/negative responses (alpha vs. not alpha). Three concepts were used: affirmation, biconditional, and conjunction/disjunction. In the simple affirmation, type of response label had no effect. In the biconditional, a complex concept in which the subsets of stimuli associated with the responses are logically identical, the use of positive/negative labels was facilitative. In the conjunction/disjunction, a complex concept in which the stimulus subsets are logically different, the use of positive/negative labels was facilitative if the simpler conjunctive subset was positive, but detrimental if the more complex disjunctive subset was positive. The results suggest that the use of positive/negative labels focuses attention on the positive stimuli and this, in turn, affects concept difficulty.

The purpose of the present study is to compare the difficulty of concept attainment under two response conditions. One condition involves neutral response labels such as alpha and beta. The other condition involves positive and negative response labels such as alpha and not alpha.

The concept attainment task requires that *S* learn to assign a set of responses to some larger set of stimuli. The kind of response label which is used does not affect the logical structure of the task, and both neutral and positive/negative response labels have been employed in previous studies. However, it is quite possible that concept attainment performance is affected by the kind of response label used. It is well known that people tend to focus on the positive subset of stimuli (Bruner, Goodnow, & Austin, 1956; Hovland & Weiss, 1953). Thus, when positive/negative labels are used, attention should be focused on the positive subset of stimuli, but when neutral labels are used attention should be divided between the two subsets. This difference in focus could affect the difficulty of concept attainment.

Peters and Denny (1971) compared learning of a conditional concept and a biconditional concept with either positive/negative or neutral labels using a rule learning procedure. They found that the conditional concept was easier with neutral labels than with positive/negative labels, but that the biconditional concept was slightly easier with positive/negative labels than with neutral labels. They examined *S*s' hypotheses and found that *S*s in the positive/negative condition tended to describe only the positive subset of stimuli, while *S*s in the neutral condition tended to describe both subsets of stimuli. Thus, it appears that the use of positive/negative labels does focus *S*'s attention on the positive subset of stimuli, but the effect this focus has on concept difficulty is dependent upon the particular concept.

The present study is designed to methodically investigate the effects of using different kinds of response labels in several different concepts, using a complete learning procedure. Three different types of concepts are used: a simple concept (one relevant attribute) in which the stimulus subsets are logically identical; a complex concept (two relevant attributes) in which the stimulus subsets are logically identical; and a complex concept (two relevant attributes) in which the stimulus subsets are logically different.

[1] This research was supported by Grant MH 14229 from the National Institute of Mental Health to Yale University.

[2] Requests for reprints should be sent to Richard L. Gottwald, who is now at Indiana University at South Bend, Department of Psychology, 1825 Northside Boulevard, South Bend, Indiana 46615.

Method

Stimuli and attributes.—The stimuli were schematic drawings of "bugs" which varied on four binary attributes: body shape, spot color, leg number, and antenna type.

The relevant attributes for all concepts were body shape and spot color. Preliminary work with direct rating and free classification tasks has shown that these are the most salient attributes of these stimuli.

Concepts and response labels.—Three different concepts were used. Examples of these three concepts with neutral response labels are shown in Table 1. In the conditions with positive/negative labels, the response assignments were identical to those in the conditions with neutral labels except that the responses were either alpha and not alpha or beta and not beta. In conditions with positive/negative labels, one subset was labeled positive for one half of the *S*s and the other subset was labeled positive for the other half of the *S*s.

The simplest concept will be termed an *affirmation*. In this concept, the subset of stimuli associated with each response is specified by the presence of one value on the single relevant attribute. For example, the alphas are round bodied and the betas are thin bodied. For one half of the *S*s, body shape was relevant and for the other half spot color was relevant. Since the subsets of stimuli are logically identical, the data from the two positive/negative conditions were pooled. There were eight *S*s in the neutral condition and a total of eight *S*s in the positive/negative condition.

The next concept will be termed a *biconditional*. Again the two subsets of stimuli are logically identical. Each subset may be described with a biconditional rule (the alphas are round and red or not round and not red, the betas are round and green or not round and not green). Since the stimulus subsets are logically identical, the data from the positive/negative conditions were pooled. There were eight *S*s in each response condition.

The third concept will be termed a *conjunction/disjunction*. In this concept, the stimulus subsets are logically different. One subset may be described with a conjunctive rule (the alphas are round and red), and the other subset may be described with an inclusive disjunctive rule (the betas are thin or green or both). Since the stimulus subsets in this concept are logically different, the data from the two positive/negative conditions were not pooled. There were eight *S*s in the neutral response condition and eight *S*s in each of the positive/negative response conditions.

In every condition the set of stimuli was arranged in such a way that each response was assigned to half of the stimuli.

Procedure.—A set of 16 stimuli was prepared for each concept. The stimuli were presented to each *S* in four random orders. The stimuli were presented in the form of slides and back projected.

The *S* was told that he was in a learning experiment and his task was to learn to discriminate between two classes of bugs which were labeled alpha and beta (or alpha and not alpha, etc., depending upon the condition). The *S* was not given any information about the nature of the possible solutions to the concept.

When a stimulus appeared, *S* responded verbally. The *E* provided immediate verbal feedback. The task was self-paced. After *S* had responded to the set of 16 stimuli, he was asked to write a rule which would discriminate between the two subsets of stimuli. Guessing was encouraged. This procedure was repeated until *S* reached a criterion of one perfect classification of the set of 16 stimuli or until *S* had responded to a total of 256 stimuli.

Subjects.—A total of 56 male and female *S*s was drawn from the psychology *S* pool at Yale University

Results

Trials to criterion.—Table 2 shows median trials to criterion (excluding the 16 criterion trials) for the various conditions. The distribution of scores was skewed so medians were used and all significance testing was done with nonparametric statistics (Mann-Whitney *U* test, unless otherwise noted).

In the affirmation concept, there was no difference between the two response conditions ($p > .50$). In the biconditional con-

TABLE 1

Attribute-Response Relationships for Three Concepts

Attributes and values		Concepts		
Shape	Color	Affirmation	Biconditional	Conjunction/disjunction
round	red	α	α	α
round	green	α	β	β
thin	red	β	β	β
thin	green	β	α	β

TABLE 2

Median Trials to Criterion

Concept	Type of response label	
	Neutral	Positive/negative
Affirmation	5.0	2.5
Biconditional	256.0	54.0
Conjunction/disjunction	26.0	
Conjunction positive		10.0
Disjunction positive		61.0

cept, the neutral response condition was significantly harder than the positive/negative condition ($p < .01$). In the neutral response condition, five of the eight *S*s failed to reach criterion, while in the positive/negative condition all *S*s reached criterion. In the conjunction/disjunction concept, the neutral condition was of intermediate difficulty. The positive/negative condition was significantly easier if the conjunctive subset was labeled positive ($p < .01$). The positive/negative condition was harder if the disjunctive subset was labeled positive ($p < .05$). Within the positive/negative condition, learning was easier when the conjunctive subset was positive than when the disjunctive subset was positive ($p < .001$).

In general, the relative order of difficulty of the three concepts was the same as that found by other investigators (e.g., Gottwald, 1971; Neisser & Weene, 1962) with the affirmation concept being the easiest and the biconditional being the hardest ($p < .05$ for every comparison). The only exception to this order was that with positive/negative labels, the biconditional concept did not differ from the conjunction/disjunction concept when the disjunctive subset was positive ($p < .40$).

Rules.—The precriterion and final rules were also examined. In the neutral response conditions, the rules described either or both subsets of stimuli. In the positive/negative conditions, most rules described only the subset associated with the positive response. This difference is seen most clearly in the final (correct) rules. In the neutral conditions, 58% of the *S*s described both subsets, while in the positive/negative conditions 75% of the *S*s described only the positive subset and no *S* described only the negative subset. This difference in descriptions is statistically significant, $\chi^2(1) = 4.20$, $p < .05$, and replicates the findings of Peters and Denny (1971).

The ways in which the stimulus subsets were described, however, were not affected by the response condition. In the affirmation concept, all final rules described the stimulus subsets using the affirmative rule. In the biconditional concept, all final rules described the subsets as a pair of conjunctions. For example, referring to Table 1, the alphas would be described as round and red or thin and green. In the conjunction/disjunction concept, all rules described the conjunctive subset with a conjunctive rule. The disjunctive subset of the conjunction/disjunction concept was described a total of 13 times. One of these rules used a disjunctive rule. The other 12 rules described the subset as an affirmation concept and a conjunction concept. For example, referring to Table 1, the betas would be described as thin or round and green.

Discussion

The primary findings of the present study are that type of response label affects concept difficulty and the nature of the effect varies with the type of concept.

In the simple affirmation concept, the type of response label had no effect. In the biconditional, a complex concept in which the stimulus subsets are logically identical, the use of positive/negative labels facilitated learning. In the conjunction/disjunction, a complex concept in which the stimulus subsets are logically different, the use of positive/negative labels was facilitative or detrimental depending upon which subset was positive.

Free recall versus paired-associate learning.—Whitman and Garner (1963) pointed out that the concept attainment task involves aspects of both paired-associate learning and free recall learning. Concept attainment is similar to paired-associate learning because some set of responses must be assigned to some set of stimuli. It is similar to free recall learning because the characteristics of the subset of stimuli associated with each response must be learned. They argued that the major part of the concept task is perceiving the relations among the stimuli within each subset and thus concept attainment is more similar to free recall learning than to paired-associate learning.

In the present study, the finding that type of response label affects difficulty is consistent with this argument. When response labels are changed, the paired-associate part of the task is not changed—each response is still assigned to eight stimuli. However, the use of positive-negative labels focuses attention on the positive subset of stimuli and this would

facilitate learning of the characteristics of these stimuli. Overall difficulty, though, is also affected by the internal structure of the positive subset. The difference in difficulty between the two positive/negative conditions in the conjunction/disjunction concept is due to the differences in internal structure between the stimulus subsets. Logically, the conjunctive subset has a simpler structure than the disjunctive subset. The *S*s' descriptions of these subsets show that the subsets are also perceived as differing in complexity. The conjunctive subset is described as one group of stimuli all of which have two attribute values in common. The disjunctive subset is described as two groups of stimuli. In one of these groups, the stimuli have one value in common, and in the other the stimuli have two values in common.

Rules.—Neisser and Weene (1962) listed 10 logically different conceptual rules which involved no more than two attributes. Each stimulus subset in the present study could have been correctly described by at least two of these rules. In a study using neutral response labels, Gottwald (1971) found that *S*s used few of these rules as their final rules, but instead described the subsets as combinations of affirmations and conjunctions. The results of the present study confirm this finding and extend it to conditions involving positive/negative labels. More generally, *S*s tend to describe the stimulus subsets in terms of one or more critical attribute values or features which are present in all stimuli within the subset. If all of the stimuli within a subset do not have features in common, the subset is broken down into groups of stimuli which do have features in common.

REFERENCES

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

GOTTWALD, R. L. Attribute-response correlations in concept attainment. *American Journal of Psychology*, 1971, in press.

HOVLAND, C. I., & WEISS, W. Transmission of information concerning concepts through positive and negative instances. *Journal of Experimental Psychology*, 1953, 45, 175–182.

NEISSER, U., & WEENE, P. Hierarchies in concept attainment. *Journal of Experimental Psychology*, 1962, 64, 640–645.

PETERS, K. G., & DENNY, J. P. Labeling and memory effects on categorizing and hypothesizing behavior for biconditional and conditional conceptual rules. *Journal of Experimental Psychology*, 1971, 87, 229–233.

WHITMAN, J. R., & GARNER, W. R. Concept learning as a function of form of internal structure. *Journal of Verbal Learning and Verbal Behavior*, 1963, 2, 195–202.

(Received March 1, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 34–42

RETROACTIVE MASKING WITHOUT SPATIAL TRANSIENTS [1]

SIDNEY STECHER [2]

Brandeis University

The threshold for the first of two 2° flashes of equal size presented to the same retinal location was determined as a function of the temporal interval between the fields. Test field duration, mask luminance, adaptation level, and mode of viewing were investigated. Masking was found to be more extensive than previously reported and central rather than peripheral factors were implicated in the effect.

Recent investigations by Westheimer (1965, 1967) and by Frumkes and Sturr (1968) showed that increases in the masking field diameter in the increment detection of a test field led to a complex function such that as the area increases the threshold first increases, then decreases, and thereafter remains constant. While these and other spatial variations in the test and background fields under increment and masking threshold detection have had extensive previous investigation (Barlow, 1958; Battersby & Wagman, 1959, 1962; Crawford, 1947; Graham & Bartlett, 1940; Ratoosh & Graham, 1951; Steinhardt, 1936), very little work has been done on the detectability of equal-size fields successively presented to the same retinal location without the presence of spatial transients between the fields.

Sperling (1965) has pointed out that spatial transients exist with unequal-size, *simultaneously* presented test and mask fields, while both spatial *and* temporal transients occur with unequal size *asynchronous* presentation of the two fields. He has suggested that discrimination in the former case is mediated by simultaneous contrast between the fields, while in the latter design (masking) the discrimination is mediated by spatial contrast *and* temporal resolution between the fields. Stecher (1968) has shown that spatial interaction between visual fields yields luminance discrimination functions different from those yielded by temporally interacting fields.

In one of an extensive series of investigations of the "Crawford effect," Battersby and Wagman (1962) determined the threshold for a 5-msec., 40-ft. test field followed after various durations by a 500-msec., variable-area masking field of 100 mL. superimposed on a 10-mL. adapting luminance under both monocular and binocular observation. They found that the test field threshold rose systematically as the mask field diameter decreased, with the largest threshold changes occurring when the test field and mask field were equal in size. The influence of these spatial factors on the test field threshold led the authors to conclude that there was "some unique process involved when the two targets are of the same diameter" and that for these equal-area, successively presented stimuli "processes of uniquely peripheral origin appeared to be involved [p. 364]," whereas for stimulus conditions in which the mask was larger than the test field, central processes were implicated.

For this special condition of monocular viewing where both fields were of equal size, the maximum threshold increase of approximately 2.5 log units occurred when the test field onset preceded the onset of the masking flash by 25 msec., with relatively little threshold change occurring beyond a temporal separation of 100 msec. Binocular observation yielded similar results, the magnitude of the threshold changes being somewhat smaller.

[1] This research was supported in part by National Institutes of Health Grant EY–00023 and by a National Institutes of Health Biomedical Institutional Support Grant to the author. The paper is partially based on a talk presented at the Eastern Psychological Association Convention held in Atlantic City, New Jersey, April 1970.

[2] Reprints should be requested from Sidney Stecher, Department of Psychology, Brandeis University, Waltham, Massachusetts 02154.

Sperling (1965) and the others cited above have demonstrated a variety of interesting relationships between the masking field parameters, adaptation luminance, and test field threshold in the simultaneous presentation (simultaneous contrast) condition. Battersby and Wagman (1962) and Crawford (1947) have demonstrated the influence of similar parameters in the asynchronous (masking) condition.

It is of importance, therefore, to determine the effects of similar parameters in the temporal interaction condition of the successive presentation of two *equal-area* fields to the same retinal location without the influence of spatial transients. The present study investigated the threshold changes induced in a test field, which preceded a masking field of equal size, as a function of: (*a*) the luminance of a constant duration masking stimulus, (*b*) the duration of the test field, (*c*) the background adaptation luminance, and (*d*) the effects of binocular and monocular presentation.

Method

Apparatus.—The apparatus consisted of a specially constructed four-channel, binocular, Maxwellian viewing system, the complete details of which have previously been published (Stecher, Sandberg, & Minsky, 1970). Briefly, the output of the system was generated by two glow modulator tubes bathed in ultraviolet light to avoid onset jitter, the outputs being optically superimposed, each eye receiving the combined flux from each of the tubes (Sylvania, R1131C). These light sources were run at a constant current of 26 ma. and could be independently controlled in duration, frequency, and interval between onsets of the flashes by a Grass physiological stimulator (S–8) pulsed through a triode-connected gating circuit. Luminance was controlled by neutral-density wedges and filters placed in an image plane in each of the channels. Test fields subtended 2°; adaptation fields subtended 12°15′. Test fields were always superimposed on the adapting field and viewing was foveal through a 2-mm. artificial pupil.

Both channels presented to each eye were optically in focus on a field lens. In front of the field lens was a field stop, with a 1.3-cm.-square opening, through which the light from each channel passed. An image of the crater was formed by the field lens in the plane of *S*'s pupil after passing through a beam splitter, being then reflected by a final mirror and passed through the 2-mm. artificial pupil. The field stop was at a distance of 38 cm. from the artificial pupil, the opening thus subtending 2° [2].

To the side of the beam splitter (through which the light from the field lens was transmitted) was a 12°15′ square adapting field. Light from this adapting field passed through neutral-density filters and was then reflected through the beam splitter into *S*'s pupil. The adapting field consisted of flashed opal glass evenly illuminated from behind by a 6–w. bulb (Sylvania).

A constant proportion of the output from each of the glow modulators was reflected onto the cathode of one of the two photocells (RCA 929). The rise and fall times for 10-msec. flashes were about 10 μsec., and a close approximation to a rectangular pulse was obtained from each output. The outputs were continuously monitored by a dual beam cathode-ray oscilloscope (CRO) throughout the course of experiments.

With each eye, *S* thus had a Maxwellian view of the opening in the field stop that was superimposed on a fixed adapting luminance. The *S* sat in the dark, and his head was held in place by a dental-impression biting board.

Centered above and below the square 2° field in each eye, two dim red points served as fixation stimuli for *S*. The *S* was asked to fixate on an imaginary vertical line between them. The fixation points were in the plane of focus of the stimulus array and were set just above *S*'s threshold by adjusting a rheostat that controlled the current passing through them. The stimulus arrangement consisted of the square 2° field presented in succession to *S*'s right eye, left eye, or both eyes. Either the first flash or the second could be set to various luminances and durations. Either of the flashes could be occluded from view to either eye. With both fields presented simultaneously (one to each eye), they appeared of equal heights and widths and were superimposed without any detectable juxtaposition of the fields.

The wedges were arranged so that different wedge positions could be attained by manual operation by *E* or *S*, or by remote-control operation by *S* alone. This was accomplished by the use of a tape reader (Ohr-tronics, 119) whose output controlled the position of a motor Bodine, KYC–22RD). The motor, in turn, governed the wedge position in each of the four channels. The output from the tape reader randomly activated a preset potentiometer. This caused the motor to stop at a position corresponding to a particular resistance setting, the latter having been set by *E* to correspond to a particular stimulus value. The apparatus was capable of presenting one of seven randomized stimuli to *S* upon command of *S*. Responses were recorded by a Gerbrands event recorder for the discrete trials used in the study. Responses of *S* could also be recorded by a chart recorder (Heath Servo-Recorder, EUW-20A) precalibrated for wedge position for continuous brightness matching tasks used in other experiments (Stecher & Sandberg, 1970; Stecher et al., 1970).

Procedure.—Before the beginning of each experimental session, *S* was dark adapted for 15 min. and then looked at the prevailing luminance of the 12°15′ adapting field for 5 more min. with either the right eye alone or with both eyes. In each of the experiments, the threshold luminance of the first flash was determined when followed after various intervals by a 10-msec. masking flash set to various luminances. The method of constant stimuli was used for all threshold determinations, *S* indicating whether the first flash was present or absent. A least-square ogive was fitted to the percentage "seen" judgments and the mean of this function was taken as the threshold estimate.

Two experiments were performed in the course of the study, the principal independent variable being the temporal separation between the two equal-area, successively presented 2° fields flashed onto the same retinal location.

In Exp. I changes in the test-field threshold as a function of the interstimulus interval (ISI) were evaluated for various durations of the test flash. The masking field luminances ranged from .1 to 1,000 mL. in log unit steps, while the duration of the mask was always kept constant at 10 msec. The test flash was set to 10, 20, 40, or 80 msec. for each of the above conditions. The ISI were always measured from the cessation of the test flash to the onset of the masking flash, and as the test flash always preceded the masking flash these intervals were designated as negative intervals in order to facilitate comparison with other studies. All test and masking flashes in Exp. I were superimposed on a .0038-mL. adapting field. Viewing was monocular with the right eye. Resting thresholds (hereafter designed as AL or RT) were determined without the presence of the masking fields and were checked each day.

In the second experiment, the threshold for the first of two 10-sec. flashes separated by various ISIs was determined when the flashes were superimposed on different adapting luminances and the viewing was either binocular or monocular. The two highest mask luminances of 100 and 1,000 mL. were used. The luminance of the adapting field was set to .0038, .038, .38, 3.8, or 38.0 mL. and was continuously present. In the monocular condition, both equal-size (2°) test fields were presented to the same retinal area in succession. In the binocular condition, the phenomenal appearance was identical to the monocular case, but the flashes were presented to corresponding retinal areas, the 10-msec. test flash going to the right eye and the 1,000- (or 100) mL., 10-sec. masking flash to the left eye.

The ISIs used were 95, 120, 150, 180, 250, and 500 msec. Each threshold value was based on 210 judgments by *S*. In each study, the appropriate stimulus parameter was randomized.

Results and Discussion

Figure 1 is a plot of the test-field threshold in log millilamberts as a function of the interstimulus intervals between the successively presented flashes. The four graphs in the figure represent the results obtained using the four different test field durations in the presence of a constant 10-msec., variable-luminance masking flash presented after the cessation of the first flash by the intervals indicated on the abscissas. The horizontal line labeled RT in each of the graphs represents the value of the 10-msec., test field resting threshold, which is the threshold value for the test field presented on the continuous, .0038-mL. adapting field without the presence of the masking luminance. The parameter of the curves within each graph is the luminance of the masking field.

With the exception of the 80–10 condition, threshold values are not shown at zero and other short ISIs. At these ISIs for the conditions investigated in the present study, there was no luminance of the test flash that made the test field detectable. The shortest ISIs at which thresholds are plotted represent the results of a preliminary experiment and are approximations to the shortest ISIs at which two flashes are detectable 100% of the time. These values are very close to those reported by Manake (1958) for comparable conditions.

The four graphs have many features in common. Generally, independent of test field duration, the functions indicate the following.

1. The log threshold remains relatively horizontal until some short ISI, at which time further decreases in ISI lead to increases in log threshold. The magnitude of the increase in threshold is between 1–2 log units at an ISI of − 100 msec. Previous investigators have indicated that such large magnitudes of threshold increase are to be expected at ISIs closer to 0 msec. and that the threshold at − 100 msec. is relatively uninfluenced by the masking flash in fields were spatial transients occur. Crawford (1947), for example, using a 12° background exposed for 524 msec. and a .5° test field exposed for 10 msec. found that the test field threshold first began to rise 100 msec. before the

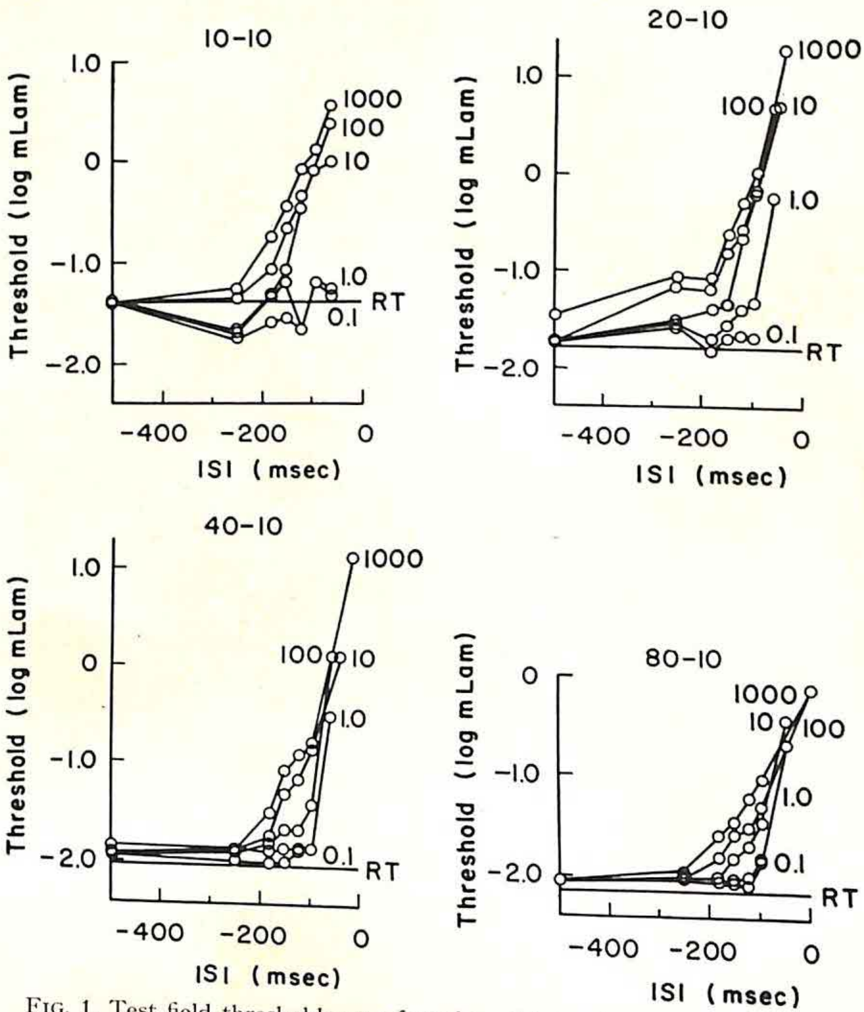

FIG. 1. Test field thresholds as a function of interstimulus interval (ISI in msec.) for four different test field durations. (The parameter of the curves is the 10-msec. masking field luminance. RT refers to resting threshold values obtained without the presence of the mask luminance. Adapting luminance is .0038 mL.)

application of the background and increased to about 10:1 at a 20-msec. separation between the fields. Beyond 100 msec., however, the functions were horizontal in this earlier study.

2. There is an ordering of the placement of the functions according to mask luminances, higher masking luminances leading to higher log-threshold values, with the lowest two mask luminances hardly influencing the test field threshold. With the 10-mL. mask luminance, however, the log threshold rapidly increases and does so in greater amounts with the 100- and 1,000-mL. mask.

3. There is an indication that the ISI at which the threshold first begins to increase is generally longer for higher mask luminances. That is, higher masking luminances yield threshold increases at longer ISIs than do lower mask luminances. Thus, with a 10-mL. mask the threshold begins to increase at about − 160 to − 180 msec.; with a 100-mL. mask at about − 220 msec. and at about − 250 msec. with the 1,000-mL. flash. Studies by Battersby and Wagman (1962), Frumkes and Sturr (1968), Crawford (1947), and others have indicated that the threshold first begins to rise between − 50- and − 100-msec. ISI

between spatially contiguous fields. Sperling (1965) using a 250-msec., 1.38° mask set to 50 mL. and a .36° test field exposed for .04 msec. has demonstrated a comparable change at zero ISI, the threshold being flat beyond − 50 msec. However, obtaining the threshold by the use of two equal-size fields presented to the same retinal location in succession demonstrates backward masking of a larger magnitude and over a more extensive temporal interval than most previously reported threshold detection studies.

The effect of changes of test field duration are also presented in Fig. 1. While the standard deviations are not presented, they were always between 15% to 30% of the threshold values. The data indicate that the test field threshold decreases at any ISI as the duration of the test field increases, and the longer the test flash duration, the shorter the ISI at which measurements were possible. Note, however, that the value of the RT also decreases as the test field duration increases, a finding in accord with the results of many previous investigators.

The extent of masking can best be indicated not by assessing the threshold but by looking at the threshold elevation, a sample of which is presented in Fig. 2. In this figure, the log of the difference between the resting threshold and the masked threshold (the latter in the presence of the masking field) is plotted as a function of the stimulus onset asynchrony (SOA). This figure indicates that the log threshold elevation is independent of test flash duration when plotted against SOA, all the different test field duration functions collapsing onto one curve. The data indicate that the threshold increase is over a range of four log units and that the threshold rise begins between − 300- and − 250-msec. SOA. Both functions indicate that the shortest *SOA* at which measurement of threshold could reliably be made for these test flash durations was − 50 mesc. This value did not change appreciably with mask luminance.

A number of previous investigators have studied the effects of background lumi-

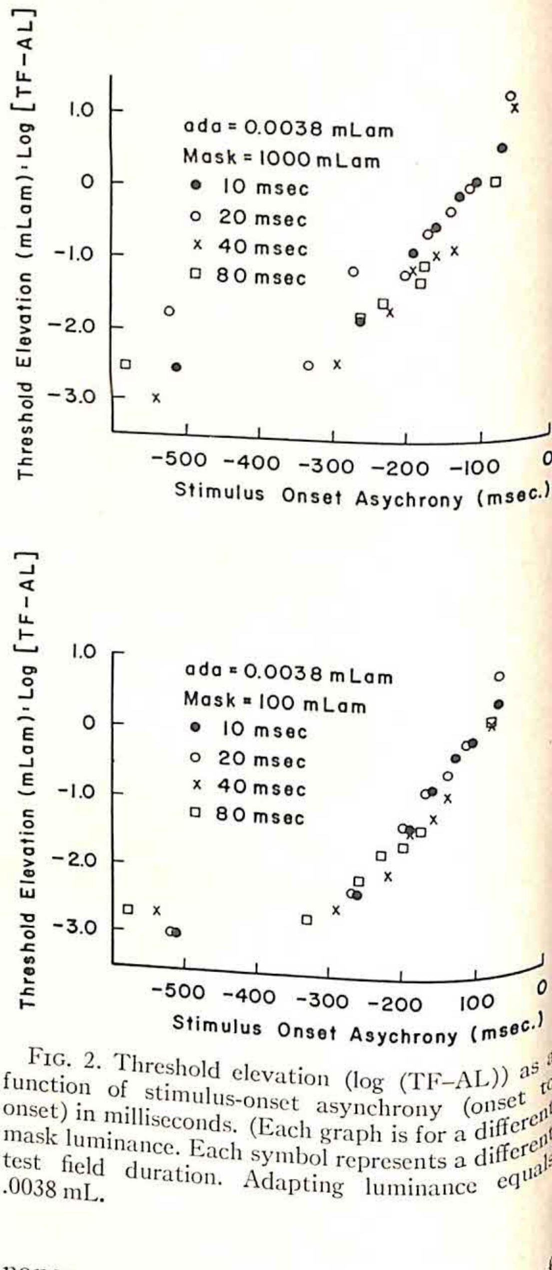

FIG. 2. Threshold elevation (log (TF–AL)) as a function of stimulus-onset asynchrony (onset to onset) in milliseconds. (Each graph is for a different mask luminance. Each symbol represents a different test field duration. Adapting luminance equals .0038 mL.

nance on threshold under a variety of parametric conditions. Sperling (1965) has suggested that the threshold elevation found at the peak of the masking function is independent of adaptation luminance. In the present study the monocular and binocular threshold for the first of two equal-size, 10-msec., successively presented flashes was obtained when these flashes were superimposed on a larger adapting field set to various luminances.

Figures 3 and 4 present the results obtained in the adaptation investigations

where the monocular threshold, in log millilamberts, is plotted as a function of the interstimulus interval. The parameter of the curves is the adaptation luminance in millilamberts. Each graph represents the threshold values obtained in the presence of a different mask luminance (1,000 and 100 mL.). The horizontal lines represent the threshold values obtained for a 10-msec. test flash superimposed on an adapting luminance without the presence of the 10-msec. masking flash. The functions indicate that the threshold (in log mL.) values obtained at all adaptation luminances tend to converge to the value of the threshold obtained at the maximum threshold values found at very short ISIs. The increase in threshold is seen to start at short ISI as the adaptation luminance increases. At an adapting luminance of .0038 mL., the threshold begins to increase at about − 250 msec. This trend is more dramatic for the 1,000-mL. mask than for the 100-mL. mask; increases in the adaptation luminance tend to shorten the interval over which the masking flash is effective. Comparison of the data with

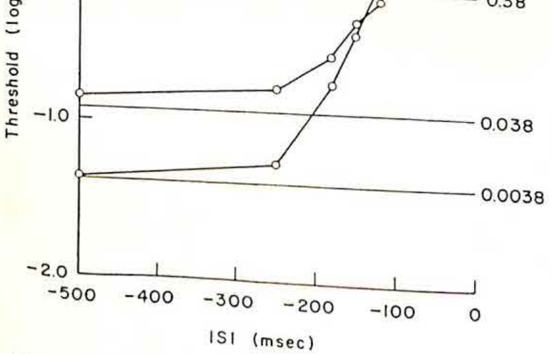

FIG. 3. Test field threshold in log mL. as a function of the ISI (msec.) for two successive 10-msec. flashes presented to the same retinal location adapted to the luminances indicated as the horizontal lines. (Mask luminance equals 1,000 mL.)

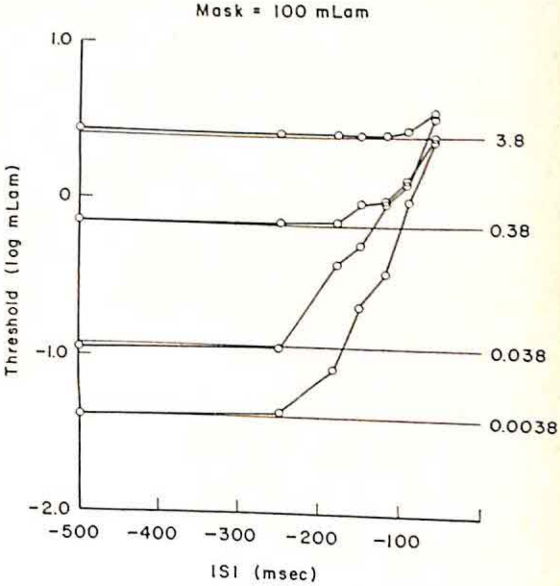

FIG. 4. Same as Fig. 3 except mask luminance equals 100 mL.

the results of Battersby and Wagman (1962), Sperling (1965), and others using comparable adaptation and masking luminance indicates the same order of magnitude of threshold increase at comparable ISI.

Thus, Battersby and Wagman (1962) using equal-size fields presented to the same retinal location in succession and superimposed on a 10-mL. adapting field found no elevation of the threshold at ISI less than − 100 msec. Sperling in the study previously cited found the threshold to be relatively uninfluenced by ISIs greater than − 50 msec. Both of these results could be contrasted with the 3.8-mL. adaptation condition which shows a rise comparable to that found in the above studies at similar ISIs. The lower adaptation functions show a greater magnitude of effect.

Caution should be used in interpreting the threshold increase from these graphs as any log threshold difference is equal to the log ratio of the threshold values or the log of the proportion threshold increase

$$\left(\log \mathrm{TD} = \log \mathrm{TF} - \log \mathrm{RT} = \log \frac{\mathrm{TF}}{\mathrm{RT}}\right).$$

A better picture of the effects of the adaptation luminance on the masking can

be obtained by plotting, at any constant ISI, the threshold elevation as a function of adapting luminance. Sperling (1965) has made a similar plot of threshold elevation as a function of mask luminance with adapting luminance as parameter and has concluded that the threshold shift is independent of the adapting luminance.

Figure 5 presents a similar plot of the present results for the two highest mask luminances at the shortest ISI at which all data points were available. The upper two curves refer to values of log threshold elevation (right ordinate). The results substantiate Sperling's findings that the *threshold elevation* is independent of these adaptation luminances. The previous figure showing the threshold values themselves gives the mistaken impression that the masking level obtained is due to the adaptation luminance. The *threshold elevation* is flat for the ISI at which the masking effect is maximum as a function of adapting luminance.

These results imply that the adaptation level sets a range over which the visual system is operative, but the masking

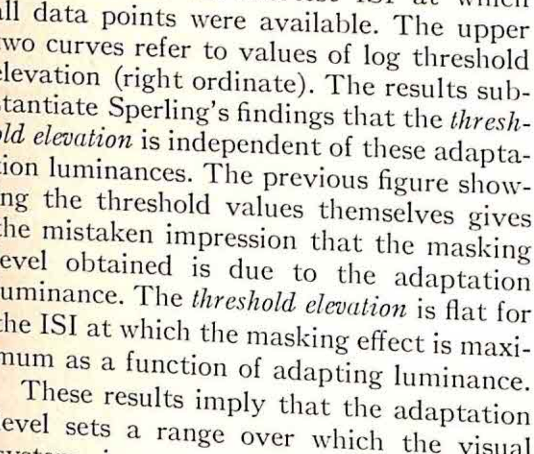

FIG. 5. Top two functions represent the log threshold elevation (log[TF − AL]) for two mask luminances for the adaptation values indicated on the abscissa. (The bottom two functions are the comparison luminances (in log mL.) found by binocular matching to appear as equally bright as the conditioning stimuli superimposed on the adapting luminances indicated on the abscissa. Open circles represent data collected in the present study; filled circles are data extrapolated from published functions of Onley and Boynton (1962).)

(threshold elevation) of a constant-duration, constant-luminance mask is independent of the sensitivity range of the eye.

Sperling (1965) has concluded that equally luminous, but unequally bright conditioning fields at zero ISI produced about equal threshold elevations. This implies that the brightness of the flashes is not the determining factor in threshold detectability. Further, since the masking effect is maximal at zero ISI, the increment threshold does not appear to be related to brightness changes, while other data indicated that luminance difference discrimination between successive flashes at superthreshold levels *are* related to aspects of apparent brightness changes (Stecher & Sandberg, 1970). The implication of this finding is that the mechanisms may indeed be different for increment threshold and suprathreshold luminance difference detectability.

In the current study, 10-msec. masking flashes of 100 and 1,000 mL. were presented on different adapting luminances ranging from .0038 to 38 mL. The two lower curves in Fig. 5. (left ordinate) represent the results of binocular brightness matches (comparison luminance) to the masking flashes superimposed on the different adapting luminances. To make these measurements *S* controlled the luminance of a 2°, 10-msec. masking field superimposed on different adapting luminances presented to the right eye. The data points are represented by open circles; the closed circles represent the values extrapolated from Onley and Boynton (1962), Fig. 2, *S* JO, in a similar study.

It is apparent from these results that in this study the masking stimuli did not change in brightness when superimposed on the different adapting luminances. In this case no dramatic changes in threshold occurred, but neither did the brightness of the constant luminance masking field change. It is interesting to note that while the threshold elevation is found to be independent of adapting luminance, as previously pointed out by Sperling (1965), the threshold elevation is roughly the same for both the 100- and 1,000-mL. masking stimuli in spite of the fact that the match-

ing luminances of the masking fields differed by about one log unit.

In the study by Battersby and Wagman (1959) cited earlier, two equal-size fields were presented in succession in the same retinal location in both monocular and binocular conditions and the monocular masking was somewhat more extensive than the binocular presentation. They presented their test and mask fields superimposed on a 10-mL. adapting luminance. The present investigation explored the effects of preadapting luminance on both binocular and monocular presentation for the successive method.

Figure 6 represents the results found for binocular and monocular conditions. The plot represented here is the log relative change in test field threshold as a function of the ISI. The axis on the right gives the percent change in threshold from rest, while the left axis yields the log ratio of test to resting threshold. This plot takes into account the differences in RT observed over the course of the experiments by evaluating relative changes from resting

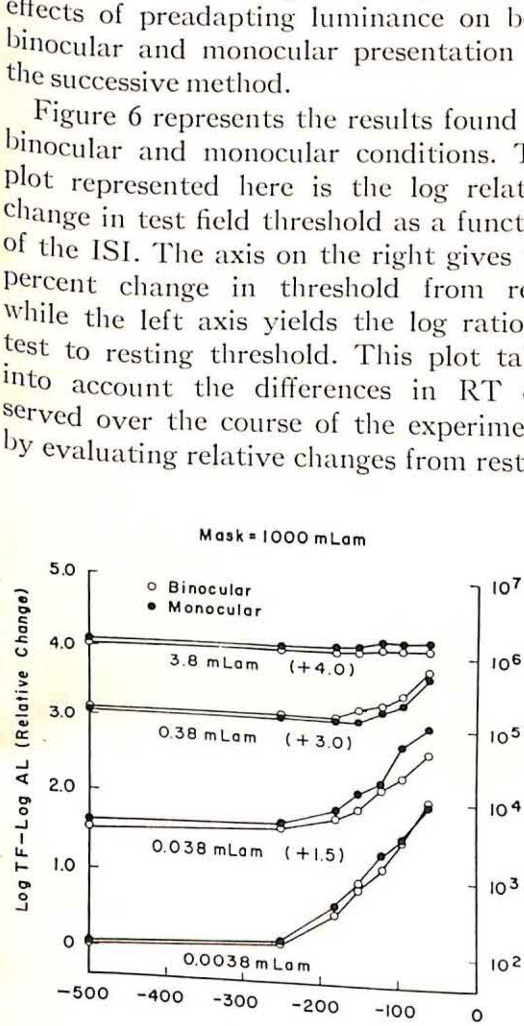

FIG. 6. Relative changes in threshold as a function of the ISI for five adapting luminances under monocular and binocular viewing. (Functions have been displaced, in log units, by amounts indicated in parentheses. All values obtained for 1,000-mL. mask luminance. Both test and mask durations equal to 10 msec.)

levels independent of the resting value found.

The three highest pairs of curves have been displaced by the amount indicated in parentheses.

The major relationships to be aware of in the figure are the lack of difference in the relative threshold changes found in the monocular and binocular conditions. This result would appear to indicate that the temporal interactions yielding the threshold effects observed in the successive method are centrally limited to the degree that binocular presentation reflects central involvement.

REFERENCES

BARLOW, H. B. Temporal and spatial summation in human vision at different background intensities. *American Journal of Physiology*, 1958, **114**, 337.

BATTERSBY, W. S., & WAGMAN, I. H. Neural limitations of visual excitability. I. The time course of monocular light adaptation. *Journal of the Optical Society of America*, 1959, **49**, 752–759.

BATTERSBY, W. S., & WAGMAN, I. H. Neural limitations of visual excitability. IV. Spatial determinants of retrochiasmal interaction. *American Journal of Physiology*, 1962, **203**, 359.

BOYNTON, R. M., & KANDEL, G. On responses in the human visual system as a function of adaptation level. *Journal of the Optical Society of America*, 1957, **47**, 275.

CRAIK, K. J. W. The effect of adaptation on subjective brightness. *American Journal of Physiology*, 1940, **94**, 232.

CRAWFORD, B. H. The effect of field size and pattern on the change of visual sensitivity with time. *Proceedings of the Royal Society, B.*, 1940, **129**, 94–106.

CRAWFORD, B. H. Visual adaptation in relation to brief conditioning stimuli. *Proceedings of the Royal Society, B.*, 1947, **134**, 283–302.

DONCHIN, E. Retroactive visual masking: Effects of test flash duration on the masking interval. *Vision Research*, 1967, **7**, 79.

FRUMKES, T. E., & STURR, J. F. Spatial and luminance factors determining visual excitability. *Journal of the Optical Society of America*, 1968, **58**, 1657.

GRAHAM, C. H., & BARTLETT, N. R. The relation of size of stimulus and intensity in the human eye. III. *Journal of Experimental Psychology*, 1940, **27**, 149–159.

MANAKE, A. Foveal discrimination measured with two successive light flashes: A psychophysical study. *Acta Ophthalmologica*, 1958, **36**, 3–11.

Onley, J. W., & Boynton, R. M. Visual responses to equally bright stimuli of unequal luminances. *Journal of the Optical Society of America*, 1962, **52**, 934.

Raab, D. Backward masking. *Psychological Bulletin*, 1963, **60**, 118–129.

Ratoosh, P., & Graham, C. H. Areal effects in foveal brightness discrimination. *Journal of Experimental Psychology*, 1951, **42**, 367–375.

Sperling, G. Temporal and spatial visual masking. I. Masking by impulse flashes. *Journal of the Optical Society of America*, 1965, **55**, 541–559.

Stecher, S. Temporal and spatial effects in luminance discrimination. *Perception & Psychophysics*, 1968, **4**, 155–158.

Stecher, S., & Sandberg, M. The relationship between binocular brightness matches and luminance difference discriminability of successively presented light flashes. *Perception & Psychophysics*, 1970, **7**, 86–90.

Stecher, S., Sandberg, M., & Minsky, P. J. Successive luminance difference thresholds and brightness as a function of the interstimulus interval and durations of successive flashes. *Perception & Psychophysics*, 1970, **7**, 79–85.

Steinhardt, J. Intensity discrimination in the human eye. *Journal of General Physiology*, 1936, **20**, 185–209.

Westheimer, G. Spatial interaction in the human retina during scotopic vision. *Journal of Physiology*, (London), 1965, **181**, 881.

Westheimer, G. Spatial interaction in human cone vision. *Journal of Physiology* (London), 1967, **190**, 139.

(Received March 4, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 43–46

RETROACTIVE EFFECTS OF PHONEMIC SIMILARITY ON SHORT-TERM RECALL OF VISUAL AND AUDITORY STIMULI [1]

PHILIP M. SALZBERG,[2] T. E. PARKS,[3] NEAL E. A. KROLL, AND STANLEY R. PARKINSON [4]

University of California, Davis

On each of a series of trials, *S*s were required to recall an alphabetical letter presented either visually or auditorily. In each case, recall followed a 12-sec. retention interval filled with auditory shadowing. Recall performance was relatively depressed when the shadow materials contained items that were phonemically similar to the memory letter but only if the memory item had been presented auditorily and, then, only if the phonemically similar items occurred very soon after the presentation of the memory letter. These results have implications for the degree to which recall of visual stimuli, under this paradigm, might be taken as a relatively pure measure of visual storage and, second, for the understanding of the effect of phonemic similarity on short-term recall of spoken stimuli.

A great deal of evidence suggests that visual stimuli are often encoded auditorily for the purpose of short-term retention (e.g., Conrad, 1964).[5] However, Kroll, Parks, Parkinson, Bieber, and Johnson (1970) argued that recall of a visually presented alphabetical letter after several seconds of "shadowing" (repeating aloud a series of spoken letter names) is based, at least in part, on storage of the visual aspects of the memory letter. That conclusion derived from the observation that under these conditions, recall of such visual items was markedly superior to recall of spoken memory letters. Thus, recall of the visual letters apparently was not typically based on an auditory memory for a covert rehearsal of the name of the memory letter. These results are understandable if, among other possibilities, a visual presentation led to a visual memory trace, that trace being resistant to interference from the subsequent auditory activities involved in shadowing.

The present research was not primarily concerned with further questioning the occurrence of persistent visual traces. Rather, the critical issue was the degree to which the shadow paradigm of Kroll et al. (1970) prevents auditory encoding of visual stimuli. To the extent that recall performance for visual stimuli under the shadow paradigm is free of the influence of auditory storage, that paradigm might provide an important research tool for the detailed exploration of visual storage.

The technique employed for detecting the presence of auditory short-term memories derives from the findings of Wickelgren (1966) that short-term recall of spoken letters was reduced when those letters were followed by a spoken list of letters whose names included vowel phonemes similar to the root vowel of the memory items. On the other hand, a visual trace would not be expected to be differentially affected by such "phonemic confusion." In fact, the finding of that effect in the recall of visual stimuli would suggest the presence of auditory encoding.

Indeed, no effect of phonemic confusion

[1] This report is based on a master's thesis submitted by the first author to Sacramento State College. The authors are indebted to George L. Parrott and Lawrence Meyers of that institution for their helpful comments. The work was supported by Grant MH 19327-01 from the National Institute of Mental Health and Grant D-262 from the University of California, Davis.

[2] Now at the University of Colorado.

[3] Requests for reprints of this article should be addressed to T. E. Parks, Department of Psychology, University of California, Davis, California 95616.

[4] Now at Arizona State University.

[5] Since current evidence does not allow a precise distinction between auditory and articulatory encoding (Wickelgren, 1969), the single term, "auditory," will be used here to encompass both possibilities.

was apparent in the visual recall data of Parkinson, Parks, and Kroll (1971, Exp. III). However, that finding, while encouraging, was inconclusive for the present purpose. In their experiment, the shadow materials that were of low similarity to the memory letters (and might therefore have favored retention) were also relatively difficult to shadow. Since this latter factor severely reduced retention of visual stimuli, it may, thereby, have obscured any tendency toward an effect of phonemic confusion. What is required is the demonstration that shadow materials which control for all variables, except phonemic similarity, do not differentially affect the recall of visual stimuli.

Method

Materials.—Each trial consisted of the presentation of 35 spoken shadow letters recorded in a female voice at a rate of 2 per sec. and the presentation of a to-be-remembered letter 5 sec. after the onset of the shadow list. Since recall occurred immediately following the last letter of the shadow list, the retention interval was approximately 12 sec.

The population of memory letters, one of which was presented on each trial, consisted of two subgroups of three phonemically similar letters each: S, F, X (having the common root vowel, ĕ) and J, K, A (root vowel, ā).

Lists of shadow letters were drawn primarily from the population C, D, I, O, P, Q, R, T, U, and Y (none of which have an ĕ or ā root vowel). However, each shadow list also included instances of two members of either subgroup of potential memory letters. For example, a particular shadow list might include instances of the letters J and K. Each such list was presented during the attempted retention of the third member of the same subgroup (in the present example, A) as well as during the attempted retention of a member of the opposite subgroup (for example, F). When a shadow list that included two members of a given subgroup was employed with the third member of that subgroup as the memory item, an instance of the High Similarity condition was formed. Presentation of the same shadow list with a memory item drawn from the opposite subgroup yielded an instance of the Low Similarity condition. Thus, contrary to Parkinson et al. (1971), identical shadow materials were employed in both levels of phonemic similarity.

Specifically, two shadow lists were constructed employing each of the six possible combinations of two members of each of the memory letter subgroups (i.e., SF, SX, FX, JK, JA, and KA) for a total of 12 master shadow lists. The 2 lists that contained the same pair of letters differed in that instances of that pair occurred either 1, 2, 3, and 4 sec. after the time of presentation of the memory letter or else occurred 7, 8, 9, and 10 sec. after that presentation. Thus, these items occurred either fairly early or relatively late in the retention interval. This variable (Early vs. Late conditions) was included as part of a second, parallel goal of the present research, a continued exploration of the effect of phonemically confusing material on short-term recall. To be precise, the data of Parkinson et al. (1971, Exp. I, 1-sec. retention condition) suggested that an effect of phonemic confusion might be most powerful very early in the retention interval.

Four copies were made of each of the master shadow list recordings for a total set of 48 lists. On two of these copies, a brief signal tone was placed on the second tape channel. This signal tone initiated the presentation of a photographic slide of the memory letter which occurred simultaneously with the eleventh letter of the shadow list.

For the High Similarity condition, the pair of potential memory letters included as shadow letters in each list dictated the corresponding memory letter for that list. For the Low Similarity condition, a member of the opposite subgroup was randomly chosen but with the restriction that all potential memory letters appear equally often.

The second two copies of each list were used to determine the retroactive effects, if any, of these shadow materials on the memory for spoken memory letters. Naturally, the absence of an effect of phonemic confusion on visual recall would only be of interest if those same lists did differentially affect the recall of spoken stimuil. The same pairings of memory letters and lists were used as in the visual memory trials, but in this case the eleventh letter of each copy was spliced out and replaced by a male-voiced recording of the memory letter. All instances of any given memory letter were copies of a single recording of that letter.

Finally, these 48-sec. generation recordings (24 for visual memory letters and 24 for auditory memory letters) were each copied twice for a total of 96 experimental trials. One copy of each second-generation list was made complete, but the second omitted all of the shadow materials that followed the presentation of the memory letter. These 0-sec. retention-interval trials were included to test for the relative identifiability of the visual and auditory memory items.

Subjects and procedure.—Each of 32 volunteer beginning psychology students participated in three experimental sessions. The first session involved 28 to 40 practice shadow lists (the number depending on the difficulty *Ss* experienced in shadowing), followed by 16 trials each including an auditory memory letter and finally 16 trials each including a visual memory letter. The next two sessions each involved 48 trials including one block of 24 randomly arranged visual memory trials and one block of 24 randomly arranged auditory memory trials, the order of these blocks being counterbalanced across *Ss*. Each block of trials was preceded by five warm-up trials in the same memory letter

modality but using as memory items letters drawn from the population of shadow letters.

Apparatus.—The shadow materials (and the auditory memory letters, when employed) were presented binaurally through headphones (Sharpe, Model HA–10A) from a Concord tape recorder (Model F98S) connected through an external amplifier (Dyna Kit, Stereo 70). The tape recorder was a modification (EDCO) which allowed a tone recorded on a separate channel of the tape to control a Kodak Carousel slide projector which presented the visual memory letters. Each such letter was projected for .2 sec. onto a ground-glass screen which was located approximately 5 ft. from the projector and 2 ft. from *S*. The intensity of that image was reduced by a $\frac{1}{16}$-in. aperture located immediately in front of the projector. As a variable of secondary interest to the present purposes, the projected size of the letters was either 1 or 4 in. (balanced across all other variables). Since this variable was quite ineffective (yielding only a mean difference of 1.3% in recall accuracy), it has been collapsed in the analyses which follow.

Results and Discussion

Mean performance under each combination of the experimental conditions is shown in Fig. 1. Over all treatments, approximately 86% of the visual items were correctly recalled after 12 sec., as compared to only 59% of the spoken memory letters, $F(1, 31) = 89.43$, $p < .01$, despite the fact that initial perception of those items was quite comparable (99% vs. 97% correct identification), as was subsequent shadow performance (81% vs. 78% correct). These findings replicated those of Kroll et al. (1970) and of Parkinson et al. (1971) and may be taken to indicate that the visual presentations led to a form of storage that was relatively immune to the general retroactive effects of the shadow activity (e.g., visual storage).

More to the present point, if that form of storage tends to be the sole basis for the recall of visual stimuli, then, as has been mentioned, it should be possible to find phonemically similar shadow materials which differentially affect the recall of spoken, but not visual, letters. The Early shadow lists satisfied this quest. For auditory trials, 66% of the items were correctly recalled when followed by Low Similarity lists, whereas only 51% of the same items were successfully recalled after the intervention of phonemically similar

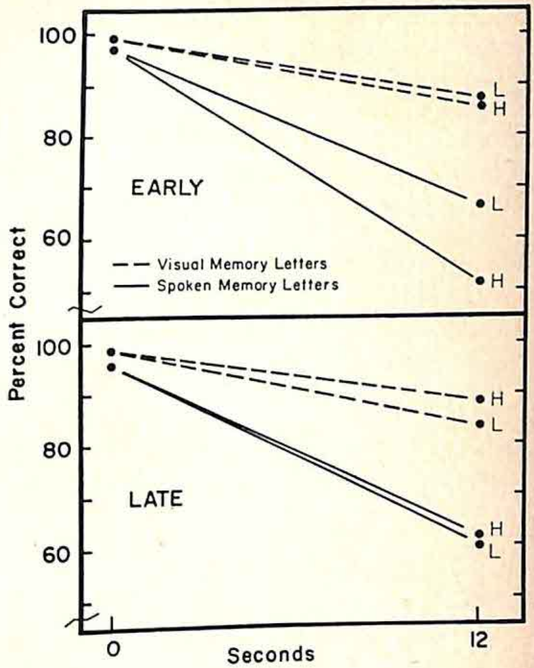

FIG. 1. The mean proportion of correct identification (0 sec.) and recall responses (after a 12-sec. retention interval) as influenced by the subsequent presentation of items that were of high (H) or low (L) phonemic similarity.

shadow letters, $t(31) = 3.258$, $p < .01$. On the other hand, the same materials produced a small, unreliable decrement of 1.6% in the recall of visual items, $t(31) = .474$. Furthermore, the Phonemic Similarity × Modality of Presentation interaction was reliable, $F(1, 31) = 6.012$, $p < .02$. This relative absence of an effect of phonemic similarity on visual memory letters would not seem to be attributable to a "ceiling effect." That is, Parkinson et al. (1971, Exp. II, III) have shown that memory traces which result from visual presentations are at least not so robust as to be impervious to all retroactive effects of subsequent auditory activity (in that study, visual recall was markedly reduced following a difficult shadow task). More probabily, the absence of any clear effect of phonemic similarity on the present visual trials reflects the fact that under this paradigm, auditory traces played an exceedingly small role in the recall of visually presented memory items.

The second purpose of the present study was the more precise delineation of the circumstances necessary for the occurrence of an effect of phonemic confusion on short-term recall of spoken stimuli. In that regard, it is quite interesting that phonemically similar items inserted late in the retention interval were, if anything, no more detrimental than Low Similarity items in the same locale (see Fig. 1). Furthermore, these High Similarity items were also ineffective as compared to the same items inserted early in the retention interval, $t\ (31) = 2.227$, $p < .05$. These findings carry with them the important suggestion that materials which are phonemically similar to a spoken memory stimulus are deleterious only if they occur very soon after memory-item presentation.

In a somewhat similar vein, several investigators (e.g., Glanzer, Gianutsos, & Dubin, 1969; Lowe & Merikle, 1971) have reported that the degree of difficulty of a task which followed a serial list of memory items differentially affected the subsequent recall of only the later items of the list. Additionally, Dillon and Reid (1969) have demonstrated that the effect of task difficulty on trigram retention is most powerful very early in the retention interval. Finally, Corman and Wickens (1968) found some tendency, though unreliable, for the retention of letters to be reduced if other letters were inserted early, rather than late, in a series of numbers that filled the retention interval. From these results, together with similar findings, one could conclude that temporal proximity to the memory stimulus may be an important ingredient of any form of retroactive interference in short-term memory for spoken stimuli. The present results extend that conclusion to the case of the retroactive effect of phonemically similar materials.

REFERENCES

CONRAD, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55**, 75–83.

CORMAN, C. D., & WICKENS, D. D. Retroactive inhibition in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 16–19.

DILLON, R. F. & REID, L. S. Short-term memory as a function of information processing during the retention interval. *Journal of Experimental Psychology*, 1969, **81**, 261–269.

GLANZER, M., GIANUTSOS, R., & DUBIN, S. The removal of items from short-term storage. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 435–447.

KROLL, N. E. A., PARKS, T., PARKINSON, S. R., BIEBER, S. L., & JOHNSON, A. L. Short-term memory while shadowing: Recall of visually and of aurally presented letters. *Journal of Experimental Psychology*, 1970, **85**, 220–224.

LOWE, D. G. & MERIKLE, P. M. Interpolation effects in short-term memory. *Psychonomic Science*, 1971, **22**, 89–91.

PARKINSON, S. R., PARKS, T., & KROLL, N. E. A. Visual and auditory short-term memory: The effects of phonemically similar auditory shadow material during the retention interval. *Journal of Experimental Psychology*, 1971, **87**, 274–280.

WICKELGREN, W. A. Phonemic similarity and interference in short-term memory for single letters. *Journal of Experimental Psychology*, 1966, **71**, 396–404.

WICKELGREN, W. A. Auditory or articulatory coding in verbal short-term memory. *Psychological Review*, 1969, **76**, 232–235.

(Received March 5, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 47–53

EFFECT OF CHOICE ON PAIRED-ASSOCIATE LEARNING[1]

LAWRENCE C. PERLMUTER[2]
Bowdoin College

RICHARD A. MONTY
United States Army Human Engineering Laboratories, Aberdeen Proving Ground, Maryland

AND

GREGORY A. KIMBLE
University of Colorado

One experiment revealed that under some conditions, A-B learning is facilitated when *S*s choose responses to be learned on a subsequent A-B list. There is also some evidence that *S*s who choose their A-B responses and are subsequently forced to learn a competing set of material (A-C) show a relatively greater disruption of learning than *S*s who did not have the opportunity to choose either A-B or A-C. All of these effects require that when *S*s choose their responses, this choosing occurs in the presence of their respective stimuli. Simply choosing responses in the absence of the stimuli produces performance which is not different from that resulting when *S*s are denied the opportunity to choose their responses.

The effect of giving *S*s the opportunity to choose the materials they wish to learn in a paired-associates (PA) paradigm is a complex, relatively unexplored area. Numerous questions arise in this context, only a few of which will be examined here.

First, it seems intuitively reasonable to assume that if *S*s are permitted to choose the response items to be learned in an A-B list, their performance should be superior to those *S*s who have been similarly exposed to alternative materials but required to learn "forced" S–R pairs (i.e., S–R pairs not of their own choosing), either because of the opportunity for *S* to form associations particularly suited to him or as a result of a general enhancement of the motivational state of the organism.

While no unambiguous tests have been made of this question, Postman's (1968) review of related work suggests the evidence to date seems to be contrary to this intuitive reasoning. For example, Underwood, Ham, and Ekstrand (1962) revealed that learning of preferred and nonpreferred materials proceeded at the same rate. In a more direct test, Brown and Read (1970) found no difference between free (choice) and yoked (forced) *S*s with either total trials or total errors to a learning criterion. However, the free *S*s were allowed to select and change their pairings throughout the learning of the list while the yoked *S*s learned only the two final lists selected by the free *S*s.

The main purpose of the present study, then, is to directly assess the effect of choosing the A–B responses prior to learning the A–B list.

Will *S*s who choose responses to be learned on A-B and who are subsequently forced to learn an A-C list perform as well on A-C as those who choose neither A-B nor A-C, but rather are forced to learn both sets of materials? If A-B bonds are stronger in the choice condition, it seems reasonable to predict they might produce greater interference later. This is the second major question examined in the present article.

Third, if *S*s are permitted to choose their responses for the A-B list, but choose in the absence of the stimulus, will the pre-

[1] This research was supported in part by Grant MH 15315 from the National Institute of Mental Health to the senior author and by Grant 8-9-150275-4005(010) from the Office of Education to the third author. Reproduction for any purpose of the United States Government is permitted. Experiment I was conducted while the senior author served as visiting professor at the Behavioral Research Laboratory of the Human Engineering Laboratories. Special thanks to Susan Luek for running *S*s of Exp. I and to Margie Rosenberger for running *S*s of Exp. II and III. Thanks also to Frank Keefe for his assistnce in data reduction.

[2] Requests for reprints should be sent to Lawrence C. Perlmuter, Department of Psychology, Bowdoin College, Brunswick, Maine 04011.

dicted enhancement of A-B learning be manifested? If enhancement was due to associations which choice *S*s formed during the choice procedure, one would predict not. On the other hand, if enhancement was due to motivational factors stemming from the opportunity to exercise choice per se, one would still expect to find enhancement.

Experiment I

In the first experiment the effects of choosing the A-B responses on learning of the A-B list and the degree or amount of transfer present from the A-B to the A-C list were examined.

Method

Subjects.—The *S*s were 40 male and female students, ages 19–25, from the University of Delaware who were paid $1.50 for their participation in the experiment.

Apparatus.—The stimuli consisted of words and paralogs typed in elite capitals on Kodak Ektagraphic Write-On slides. These slides were projected with a Kodak 850 Carousel projector programmed with a Hunter Model 111C Silenced decade interval timer at a viewing distance of approximately 1.8 m. The five-letter CVCVC words and paralogs were taken from a list generated by Andreas (1967). Andreas' entire list of 75 items was rank ordered along the meaningfulness dimension. The first 35 items were then designated as high-meaning stimuli and the last 35 items as low-meaning stimuli. The remaining 5 items were discarded. The selected words were then divided into 10 groups of 7 items each, again on the basis of rank. That is, Items 1–7 constituted the first group, Items 8–14 constituted the second group, etc. Within each group the middle ranking word was designated as the stimulus and the word immediately preceding it in rank was designated as the A-C response word for that set. The remaining 5 words served as the 5 *potential response words* on the A-B list. Thus, within any given set of 7 items, meaningfulness was reasonably homogeneous.

Procedure.—The *S*s were randomly divided into two groups of 20 each. The *Choice group* was first presented with 10 slides, 1 at a time. Each slide contained a stimulus word centered on the left side and five potential response words listed vertically on the right. The *S* was instructed to read the stimulus and five potential response words aloud and then to select one word to serve as a response word to that particular stimulus in a subsequent PA learning task. Each *S* went through each of the 10 slides at his own rate. On the average, the selection took approximately 15 sec. per slide.

Following selection of the response items, *S*s were then given 10 PA learning trials with their individually selected S–R pairs. The stimulus slide was shown for approximately 3.8 sec. and was followed by a second slide displaying the chosen S–R pair for approximately 3.8 sec. The interval between slides was approximately .9 sec. The *S* was instructed to say the appropriate response word aloud upon presentation of each stimulus slide. The *E* recorded *S*'s verbal responses.

There were 10 presentations of the 10-word list in three different random orders. Each order was presented on every third trial and each *S* viewed all items in the same order. A brief period of approximately 2 sec. followed every third presentation of the list to enable *E* to return the slide tray to its starting position.

This original or A-B list was then followed by 10 PA learning trials with pairs comprised of stimuli identical to those used on the A-B list and the *E*-selected responses indicated above. This A-C list followed the A-B list after a delay of approximately 50 sec., necessary to change slide trays on the projector. The *S*s were informed that they would be required to learn a new set of response words to the original stimulus words.

The *Force group* was treated identically to the Choice group except that although they read the stimulus and potential response words aloud, they were not given the opportunity to choose their own responses. Rather, immediately following their reading, they were assigned the response word chosen by the previously tested Choice *S* while each slide was visible to him.

Results and Discussion

The number of words correct per trial within each level of meaning served as the measure of performance on the A-B trials. Thus, the highest score obtained on a given trial was five at each level of meaning. A one-between-variable and two-within-variable analysis of variance (Butler, Kamlet, & Monty, 1969) was performed on these data to determine if there were any differences in learning of the A-B list as a function of groups (Choice vs. Force). Trials and meaning (high vs. low) were included as the within variables. Only the main effects of meaning, F (1, 38) = 254.88, $p < .001$ (with a mean number correct of 4.30 for high-meaning stimuli and 2.66 for low-meaning stimuli), and trials, F (9, 342) = 305.83, $p < .001$, and the Meaning × Trial interaction, F (9, 342) = 28.77, $p < .001$, reached statistical significance. The data underlying this interaction simply indicated that high-meaning stimuli were

learned at a faster rate than low-meaning stimuli. Groups and all interactions with groups failed to reach significance at the .05 level of confidence, which suggests that allowing *S*s to choose their own response words had no beneficial effect upon learning, a result in general agreement with those reviewed by Postman (1968). To determine if *S*s who had chosen their own materials on the A-B list would show an initial retardation in learning on the A-C list, an identical analysis of variance was performed on Trials 2–10 of the A-C list (the first trial was omitted from the analysis as all scores were zero).

As in the previous analysis, meaning, $F\ (1,\ 38) = 128.17,\ p < .001$, trials, $F\ (8, 304) = 235.57, p < .001$, and the Meaning $\times$ Trials interaction, $F\ (8, 304) = 20.67,\ p < .001$, reached statistical significance. Again, high-meaning stimuli were learned at a faster rate than low-meaning stimuli. Of greater interest, however, is the Groups $\times$ Trials interaction, $F\ (8,\ 304) = 2.03,\ p < .05$. Examination of the data underlying this interaction (Table 1) and a subsequent t test revealed that the effect can be accounted for on the basis of the superior performance of the Force group on the A-C list on Trial 2, $t(38) = 2.56, p < .02$. The superiority disappears on Trial 3 and on subsequent trials. This suggests, then, that following the learning of the A-B list, *S*s who chose their own response words showed an initial retardation in learning relative to *S*s who were not given the opportunity to choose.

The reason for this result requires further investigation. An investigation was made of A-B intrusions on the A-C trials. However, the number of intrusions was very small and did not reveal a difference between the Choice and Force groups. An additional analysis was made of the actual words chosen by the Choice and Force groups. In this examination, A-B performance was plotted as a function of frequency of word selection. Out of the possible 50 response words, 46 were chosen at least once. Some words were chosen by only a single *S*, while others were chosen by as many as 13 *S*s. Performance of these groups as a function of word-selection frequency suggested that for low-meaning words there was some tendency for the Choice group to show superior performance for the infrequently chosen words. However, for the high-meaning words this result was not evidenced. In general, this analysis did not appear to be productive for furthering the understanding of the effect of choice on learning. Had these analyses shown that for infrequently chosen words the Choice group was superior to the Force group, there might have been some evidence for the role of association in accounting for the observed choice–force performance difference on the A-B trials.

In summary then, the first experiment suggests that if a group of *S*s is forced to learn a set of materials (A-B) and then is forced to learn a competing set of materials (A-C), performance is disrupted less than if *S*s were able to initially choose A-B and then were subsequently forced to acquire a competing set of associates (A-C). This suggests that the A-B associative bonds were stronger in the Choice condition, thus producing more interference during the A-C trial. It is difficult, however, to speculate as to why this greater strength was not in evidence on the A–B trials.

TABLE 1

MEAN NUMBER OF CORRECT RESPONSES ON THE A-C TRIALS FOR THE CHOICE AND FORCE GROUPS

Group	Trial								
	2	3	4	5	6	7	8	9	10
Choice	1.08	2.45	3.05	3.55	3.72	4.05	4.28	4.58	4.68
Force	1.65	2.55	2.92	3.58	3.90	4.10	4.30	4.50	4.78

Note.—The first trial was omitted from the analysis as all scores were zero.

One possibility is that the task was simply too easy and, hence, insensitive to any beneficial results of choice which might have occurred. Although there was a tendency for the Choice group to show superior performance to that of the Force group on the A-B trials, this difference was not significant for either the high-meaning or low-meaning materials. However, the magnitude of the difference did appear larger for the materials which were more difficult to learn (i.e., low meaning). For the high-meaning materials, acquisition was very rapid, and both groups reached near perfect performance in relatively few trials. Hence, Exp. II was conducted in order to determine whether a more stringent set of conditions would allow the anticipated superiority of the Choice group to become manifest.

Experiment II

Several possible ways of increasing task difficulty were considered such as lengthening the word list, decreasing the meaningfulness of the list by including only low meaning materials, or increasing rate of presentation. The latter technique seemed to offer several advantages. First, insofar as stimulus materials are held constant across experiments, interpretation of results becomes somewhat clearer. That is, if the materials were changed, it would be impossible to assess whether any observed differences between experiments were actually due to increased difficulty or whether they were simply attributable to some peculiar aspects of the new lists. Second, increasing the rate allowed the meaningfulness variable to be preserved. Meaningfulness is also a dimension of difficulty and its inclusion allows for the possibility of assessing interaction effects. Thus, the presentation rate of material was increased from 3.8 sec. to 2 sec. during both the A-B and the A-C trials. A 2-sec. presentation rate is also one which is more commonly employed in PA learning studies.

Method

Subjects.—The *S*s were 40 male and female undergraduate students from the University of Delaware who were paid \$1.50 for their participation in the experiment.

Apparatus and procedure.—The apparatus, materials, and general procedures were identical to those employed in Exp. I, with the exception that the stimulus slides and S–R pair slides were shown for 2 sec.

Results and Discussion

As in the previous experiment, the number of words correct per trial within each level of meaning served as the basic measure of performance. Thus, the highest score obtained on a given trial was five at each level of meaning.

A one-between-variable and two-within-variable analysis of variance was applied to the A-B scores. As in the previous experiment, the main effects for trials, F (9, 342) = 108.28, $p < .001$, and meaning, F (1, 38) = 314.85, $p < .001$, reached statistical significance (with mean number correct of 3.74 and 1.68 for the high- and low-meaning material, respectively). In addition, the Groups $\times$ Meaning $\times$ Trials interaction, F (9, 342) = 4.34, $p < .001$, also attained significance. The data underlying this latter effect are shown in Fig. 1.

It can be seen that with high-meaning S–R pairs Choice group performed somewhat better than the Force group on early trials while performance tended to equalize on the later trials, presumably due to a ceiling effect. By contrast with low-meaning S–R pairs, the superiority of the Choice group over the Force group persisted even with practice. In contrast to Exp. I then, at a faster rate of presentation, *S*s who chose the materials to be remembered performed significantly better than *S*s who did not exercise choice. Further, with low-meaning materials the advantage of having the opportunity to exercise choice appears even greater. In view of the findings of Postman (1968), this finding clearly warrants further study, especially with a longer, high-meaning list so that the ceiling effect does not become a consideration.

As in the previous experiment, an identical analysis of variance was performed on Trials 2–10 of the A-C list. As in the previous analysis, meaning reached signifi-

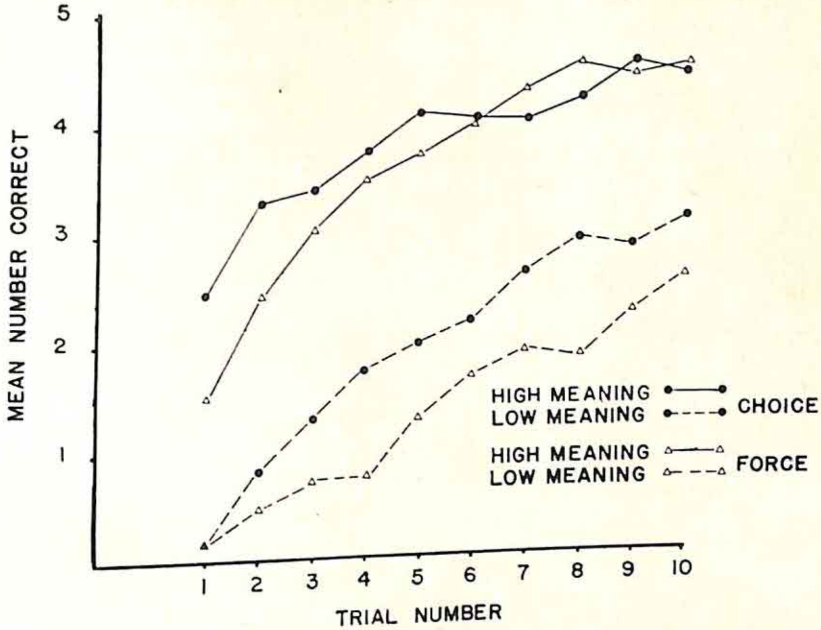

FIG. 1. Mean number of correct responses on the A-B trials as a function of groups and level of meaning (Exp. II).

cance, $F(1, 38) = 297.92$, $p < .001$, with a mean number correct of 3.58 for high-meaning stimuli and 1.66 for low-meaning stimuli, respectively. Similarly, trials reached significance, $F(8, 304) = 129.02$, $p < .001$, simply indicating that performance improved with trials. Finally, the Trials × Meaning interaction also attained significance, $F(8, 304) = 5.27$, $p < .001$, indicating that high-meaning stimuli were learned at a faster rate than low-meaning stimuli. In contrast to Exp. I, however, groups (mean number correct of 2.63 and 2.60 for the Choice vs. Force *S*s, respectively) and all interactions with groups failed to reach significance at the .05 level of confidence.

In summary, then, the increase in the rate of presentation introduced in Exp. II brought about differences in the ability of Choice versus Force *S*s to learn the initial S–R lists and also tended to accentuate the importance of the meaningfulness dimension. In contrast to Exp. I, however, no differences between groups were noted on the A-C trials. When evaluating A-C performance following A-B learning, however, the evidence generally suggests that A-C performance is dependent upon the number of A-B trials (Hall, 1971, p. 475). In this experiment, although the number of A-B trials was identical for both groups, performance on these trials for these groups was reliably different. Hence, it is difficult to directly evaluate the A-C performance of these groups without consideration of performance on the prior A-B list. A difference score analysis (i.e., A-B minus A-C) suggested that having had the opportunity to choose materials tended to cause a greater disruption of performance than having learned materials not of one's choosing; however, there are also serious objections to the use of difference scores in this paradigm so this type of analysis must be treated with caution.

EXPERIMENT III

Are the differences in the learning and transfer noted in the above experiments due to associations which the Choice *S*s have the opportunity to form during the choice procedure or simply due to the fact that they had the opportunity to exercise choice per se? Experiment III was designed to

answer these questions by requiring *S*s to select response words in the absence of stimulus words.

Method

Subjects.—The *S*s were 40 male and female undergraduate students from the University of Delaware who were paid $1.50 for their participation.

Apparatus and procedure.—The apparatus, materials, and general procedure were identical to those employed in Exp. I, with the following exceptions: (*a*) When the potential response words were shown to *S*s, they were shown in the absence of the stimulus words. Thus, the Choice group selected their response words in the absence of the stimulus words, and the Force group was required to learn these selected materials. (*b*) As in Exp. II, the stimulus slide and S–R pair slides were shown for only 2 sec., rather than for 3.8 sec.

Results and Discussion

As in the previous experiments, the number of words correct per trial within each level of meaning served as the basic measure of performance on the A-B trials. A one-between-variable and two-within-variables analysis (Butler et al., 1969) was performed to determine if there were any differences in learning of the A-B list as a function of groups (Choice vs. Force). Trials and meaning were included as within variables. The first trial was excluded from the analysis as all scores were zero.

Only the main effects for meaning, F $(1, 38) = 203.13$, $p < .001$ (with a mean number correct of 3.49 for the high-meaning stimuli vs. 1.46 for the low-meaning stimuli) and trials, F $(8, 304) = 91.05$, $p < .001$, and the Meaning × Trials interaction, F $(8, 304) = 6.77$, $p < .001$, reached statistical significance. As in the previous experiment, the data underlying this latter effect indicated that the high-meaning pairs were learned at a faster rate than the low-meaning pairs.

In contrast to Exp. II, groups (a mean number correct of 2.58 for Choice *S*s and 2.38 for Force *S*s) and all interactions with groups failed to reach statistical significance at the .05 level of confidence, which suggests that the superiority of the Choice group over the Force group noted in Exp. II can be attributed to the fact that *S*s had the opportunity to select response words which they could easily associate with the stimulus words.

As in the previous experiment, an analysis of variance was performed on Trials 2–10 of the A-C list. Again, meaning reached significance, F $(1, 38) = 222.83$, $p < .001$ (with a mean number correct of 3.82 and 1.74 for high-meaning vs. low-meaning stimuli, respectively) as did trials, F $(8, 304) = 117.30$, $p < .001$, and the Trials × Meaning interaction, F $(8, 304) = 7.96$, $p < .001$. This latter effect again indicated that high-meaning stimuli were learned at a faster rate than low-meaning stimuli. Groups and all interactions with groups failed to reach significance at the .05 level of confidence.

In summary, then, the failure to observe differences in performance as a function of the Choice versus Force conditions during the A-B trials suggests that the superiority of the Choice group over the Force group noted in Exp. II can be attributed to the fact that *S*s had the opportunity to select response words which they could easily associate with stimulus words and not simply to having had the opportunity to exercise choice. Similarly, the failure to observe differences in performance between groups on the A-C trials suggests that the interference noted in Exp. I also seems attributable to the opportunity to form bonds of one's own choosing and not to choice per se.

CONCLUSION

In conclusion, then, the experiments reported here have revealed several interesting points. First, as demonstrated by Exp. II, *S*s who have the opportunity to choose their own responses in an S–R paradigm may learn faster than *S*s who do not exercise choice. By contrasting this result with that of Exp. I, however, it appears that meaningfulness and task difficulty (rate of presentation) are both very important parameters. Further research will have to be conducted to obtain a fuller appreciation of the phenomena. The theoretical reason for this effect is not now clearly evident, however; one possibility is that *S*s who choose their response associates may

be generally motivated by this procedure, with the result that performance is superior to that of *S*s who have not been allowed to choose.

Second, Exp. I suggests that *S*s who have had the opportunity to choose their responses in an S–R paradigm and are subsequently forced to learn a competing set of materials, show a greater disruption of learning of these new materials than *S*s who did not have the opportunity to choose. The results from Exp. II are less clear unless one is willing to accept the dangers inherent in interpretation of the difference scores. If *S*s are motivated by the choice procedure, it is possible that this increased motivation hinders the acquisition of new competing materials, either because the original A-B habits are stronger in the Choice group, or perhaps because the common force procedure (A-C) serves to differentially disturb *S*s who had previously chosen but subsequently are forced on A-C.

Finally, all of the above effects would appear to be a function of the opportunity of *S*s to form S–R associations during the choice procedure. When this opportunity is removed, Choice *S*s performed nearly identically to Force *S*s. The loss of superiority on A-B by the Choice group in Exp. III suggests that in order for the presumed motivational effect to occur, it is necessary that the *S*s be permitted to choose the actual S–R units to be learned. Simply choosing may not be sufficient for the operation of this suggested motivational mechanism. Obviously, this explanation and the burden placed upon an unspecified motivational system must be seen only as a tentative suggestion and one which must await further test. One such test which has been suggested is to measure response latency during the A-B and A-C trials as a possible indicator of enhanced motivation.

REFERENCES

Andreas, B. G. Continued word association scaling of stimulus meaningfulness (*m*) and common response probabilities (*p*) for standardized verbal items. Paper presented at the meeting of the Eastern Psychological Association, Boston, April 1967.

Brown, S. C., & Read, J. D. Associative organization in paried-associate learning and transfer. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 317–326.

Butler, D. H., Kamlet, A. S., & Monty, R. A. A multi-purpose analysis of variance FORTRAN IV computer program. *Psychonomic Monograph Supplements*, 1969, 2, 301–319.

Hall, J. F. *Verbal learning and retention*. Philadelphia: J. B. Lippincott, 1971.

Postman, L. Association and performance in the analysis of verbal learning. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

Underwood, B. J., Ham, M., & Ekstrand, B. Cue selection in paired-associate learning. *Journal of Experimental Psychology*, 1962, 64, 405–409.

(Received March 8, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 54–58

EFFECTS OF REPEATED TESTS ON RECOGNITION TIME FOR INFORMATION IN LONG-TERM MEMORY

IRA FISCHLER[1] AND JAMES F. JUOLA

Stanford University

The *S*s memorized a list of 24 words and were tested the following day for recognition time in a two-choice task for list (target) and nonlist (distractor) words. The effects of repeating targets and distractors on response latency were measured. Repetitions produced shorter latencies and fewer errors for target words but longer latencies and more errors for distractors. These findings were discussed in terms of a model in which *S* judges the familiarity of the test word and on this basis decides whether to respond immediately or to execute a search of the memorized list before responding.

In a study of short-term recognition memory, Sternberg (1966) has shown that a serial search, or scan, of a target set of from one to six digits may take place when *S* decides whether or not a test digit is a member of the memorized list. This scanning process was inferred from the fact that for both positive and negative responses, latency was a linearly increasing function of the number of target-set digits.

A serial search process would seem to be inadequate for most long-term recognition tasks in which the "memory set" might consist of hundreds of items learned over a long period of time. An exhaustive serial search of all possible storage locations prior to the recognition decision seems improbable due to the speed of such recognition. To investigate the search processes involved in long-term memory, Juola, Fischler, Wood, and Atkinson (1971) employed the Sternberg (1966) paradigm in a long-term design. In this case, *S*s were tested for recognition latency to words, with half of the presented words taken from a list which had been memorized the evening before the test session. The memorized word list became the "target set" for the recognition task: target words were to be judged as positive; any other words presented were "distractors" and were to be called negative.

In the Shepard and Teghtsoonian (1961) paradigm, *S* is shown a continuous series of items, and he is to recognize repeated items as old ones. In such a design a repetition is by definition a positive item. Using this procedure with words as stimuli, Hintzman (1969) found recognition latency to decrease over two repetitions of an item. Further, in all cases, positive response latency was faster than negative latency.

Using a discrete target list, Juola et al. (1971) also found positive latency to decrease when a target word was repeated. On initial presentations, however, mean latency of responses to target words was actually greater than response latency for distractors. Since the use of a memorized list makes it possible to repeat distractor items as well as targets, Juola et al. also presented some of the distractor words a second and third time. This produced slightly *longer* latencies than those for unrepeated distractors, but the effect was not significant.

The present study was designed specifically to examine the effects of repeating target and distractor words when the target list was in long-term memory. To avoid confounding repetition effects with any decrease in mean latency which might result from practice, the test session was divided into four blocks of trials. One-fourth of the target and distractor words appeared once in each of the four blocks; another quarter were shown once in each of the last three blocks; another in the last two; and the remainder appeared only in the final block. Within each block the test items were presented in a random order. Thus mean response latency could be com-

[1] Requests for reprints should be sent to Ira Fischler, Department of Psychology, Stanford University, Stanford, California 94305.

pared across presentation number within each block.

Method

Subjects.—Twenty right-handed Stanford University undergraduates served as *S*s.

Stimulus lists.—Forty-eight one-syllable nouns having a high natural frequency of occurrence (more than 50 per million according to Thorndike and Lorge) were chosen as stimuli. Synonyms and homophones were not used, nor were words that were visually very similar (e.g., LAMP and LUMP). Word length varied from four to six letters.

From this pool of 48 words, 24 were chosen randomly to be a target list for a given *S*. The remaining 24 items served as distractor words for that *S*. For a second *S*, these distractors became the target list, and vice versa. This allowed each word to appear equally often as a target and as a distractor. This drawing-and-matching procedure was repeated 10 times, giving a total of 20 target lists with corresponding distractor words.

Apparatus.—Each of the 48 words was typed in large capital letters on a white 5 × 8 in. index card with an IBM Executive typewriter. During the test session, each word was shown singly in an Iconix tachistoscope and exposure box (System 153). The words appeared in the lower center of the visual field and subtended a horizontal visual angle of less than 2°. Between stimulus exposures, the display field was dark, except for a small circle of light (1.4 ftl.) near its center. Display brightness averaged 39 ftl., with the word appearing immediately below the central circle.

On a table to the right of *S*, three telegraph keys were arranged along an arc, with each key separated from the adjacent key(s) by about 3 cm. The *S* could comfortably rest his right forefinger on the center key between trials and could make a short, natural movement to the left or right to strike either of the two response keys. For half of the *S*s, randomly chosen, the left key was pressed to signify a positive recognition response; the right key, to indicate that the word was not on the target list. These conditions were reversed for the other *S*s. A button held by *S* in his left hand was used to initiate display exposure.

Procedure.—Each *S* was contacted by phone on the evening before the experimental session and given one of the target lists to memorize. They were told (*a*) to read through the list repeatedly, trying to recall the words in their correct serial order after each reading and (*b*) to continue until they had made one completely correct recall. It was suggested that half an hour would probably be sufficient for this.

At the start of the experimental session, it was explained that *E* was investigating how rapidly people could decide whether or not a given word was a member of the target list. The *S* was then given a sheet of paper on which his target list words were typed in a single central column in a style identical to that used on the 5 × 8 in. index cards and in the same serial order as they had been read to *S* the previous evening. One-minute study sessions were then alternated with written serial recall tests. All *S*s satisfied the same criterion by correctly recalling their target lists on the first two test trials.

The *S* was then seated comfortably in front of the tachistoscope and was told that on each trial (*a*) *S* would press a start button in his left hand; (*b*) ½ sec. later, the display would be illuminated for .5 sec., lettering *S* see the word; (*c*) *S* would then decide whether or not the word was in the list and press the appropriate key; and (*d*) *E* would record the response latency and change the stimulus cards, (with an ITI of about 10 sec). Nothing was said if *S* was correct, but he was informed of errors.

After a reminder to respond as quickly but as accurately as possible, the testing began. There was a total of 120 trials, half of them positive and half negative. The four blocks of trials were presented as a single continuous series. Six target words and six distractors were presented once in all four blocks; another six of each type were presented in the last three blocks; another six in the last two blocks; and the remaining six only in Block 4. Each block, then, repeated all the test items of the previous block, along with 12 new items appearing for the first time. This resulted in block sizes of 12, 24, 36, and 48 trials for Blocks 1–4, respectively. Order of presentation was randomized within each block, and the assignment of targets and distractors to presentation conditions was also randomized for each *S*.

Results

The overall error rate was 4.0%. In the analysis of the latencies, the error scores were omitted. For each group of six scores that an *S* produced on a presentation number in a given block, a median score was obtained. A mean latency was then obtained across *S*s for each type of trial. These data are presented in the upper portion of Fig. 1. For each block, mean latency is plotted as a function of presentation number for targets versus distractors. In each block, positive latency is greater than negative latency for targets and distractors presented for the first time. But positive latency decreases with presentation number, while negative latency increases.

Separate analyses of variance were performed for each of the curves in Fig. 1, using the medians as single scores and the *S*-presentation interaction as the error term. The decrease in positive latency with presentation number was highly significant

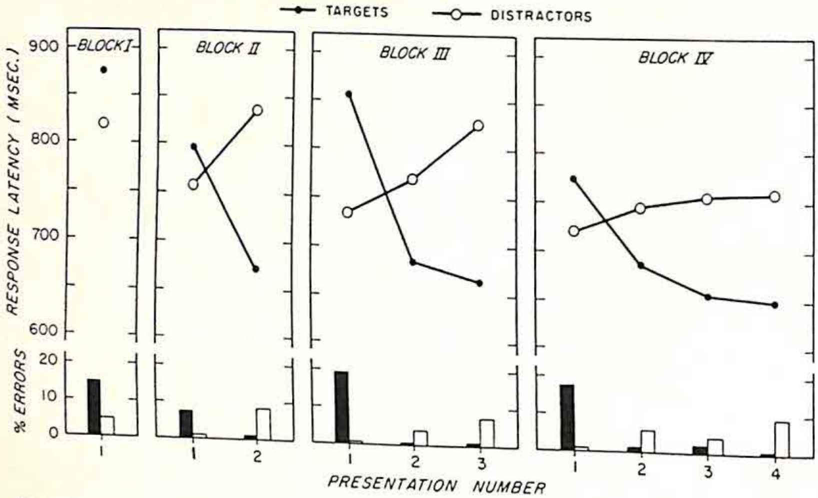

FIG. 1. Mean response latency and error frequency as functions of presentation number for targets and distractors, in each of four blocks.

in all three blocks (2: $F(1, 19) = 63.59$; 3: $F(2, 38) = 23.54$; 4: $F(3, 57) = 20.79$; all ps $< .001$). The increase in negative latency was significant in Blocks 2 and 3 (2: $F(1, 19) = 12.58$; $p < .005$; 3: $F(2, 38) = 7.32$, $p < .005$), but it was not so in Block 4 ($F(3, 57) = 1.76$, $p > .10$). In each of the three blocks, the effect of presentation number is somewhat smaller for negative than positive latency. (A Wilcoxin signed-ranks test was used, comparing the first vs. the mean of later presentations, for targets vs. distractors. The interaction was significant in each block (2: $T = 55$, $p < .05$; 3: $T = 39$, $p < .01$; 4: $T = 36$, $p < .005$.)

In the lower portion of Fig. 1, the error frequencies are presented. The solid bars are errors to targets, the open ones, to distractors. In each block, the error frequency mirrors the latency data: target words produce most errors on the initial presentation, while distractors produce most on later presentations.

The effect of serial position on response latency for target items was examined by dividing each target list into quarters and calculating a mean score for each type of trial in each quarter of the list. No trend in response latency with serial position was observed.

Discussion

Repetition has a clearly differential effect on response latency to target words as opposed to distractor words. An *S* is faster at identifying a target word, but tends to be slower in rejecting a distractor, if either had appeared before. Repetitions also produce fewer errors for targets, but more errors for distractors.

This type of repetition effect is of a different order than that found in short-term choice reaction time studies (e.g., Bertelson, 1963). In the latter, recognition latency to a stimulus was shorter if it had recently appeared before in the test series. Two general explanations have been offered for this: in one, a decaying visual trace of the first presentation is potent enough to facilitate recognition at the second presentation. In the other, an item just presented has a favored position in any search of the short-term store (cf. Atkinson & Shiffrin, 1968). These processes, however, are probably not effective over the long lags involved in the present design, where ITI was about 10 sec. and the average number of intervening items between repetitions was 30. Moreover, neither process predicts an effect of repetition that is opposite for target and distractor words.

Juola et al. (1971) modified a signal-detection model of recognition which had been presented earlier by Parks (1966) and developed by Kintsch (1967). The model assumes that each presented word has associated with it a "familiarity" value that can be represented as a point on a continuous scale. The Juola et al. (1971) model is presented schematically in

Fig. 2. In this version, if the familiarity value of an item falls above some criterion c_H, S immediately makes a positive response. Below a lower criterion c_L, he makes a negative response. Between c_L and c_H, S may retrieve portions of the target list and scan them in search of the test word; this extended search process delays the response decision, resulting in a longer latency. This model has been used to provide quantitative predictions for a variety of recognition experiments (see Atkinson & Juola, 1971).

The present results can be interpreted within the framework of this model as follows: There are assumed to be two density functions which reflect the probabilities of obtaining a particular familiarity value when S is tested with a target or with a distractor word. At the start of the test session, the expected familiarity value for target words is greater than for distractors (Fig. 2A); however, the mean of the target distribution is closer to the center of the search region than that of the distractor distribution. This means that more targets than distractors will have familiarity values falling in the search area between c_L and c_H. Positive latency, then, would include a greater number of trials on which a search of the memorized list takes place, resulting in a higher mean latency. Also, the number of target words with expected familiarity falling in the "fast no" region below c_L would be greater than the number of distractors with familiarity in the "fast yes" region above c_H. This would produce more errors to target words than to distractors.

The effect of a presentation would be to increase the expected familiarity for both a target and a distractor word. This is shown in Fig. 2B in which the familiarity distributions for previously presented target and distractor items are shifted upward. As distractors become more familiar, they are more likely to be judged between c_L and c_H. Since this means an extended search of memory takes place on a greater number of trials, mean response latency to repeated distractors will be greater than latency to initially presented nonlist words. In addition, the upper tail of the distractor distribution moves into the "fast yes" region above c_H, producing more false positive responses. The opposite of these effects occurs for target items, as the target distribution moves away from the area beteen c_L and c_H. With fewer test trials resulting in a search of the target list, mean positive latency is reduced. Errors also decrease as the lower tail of the distribution moves out of the "fast no" region below c_L.

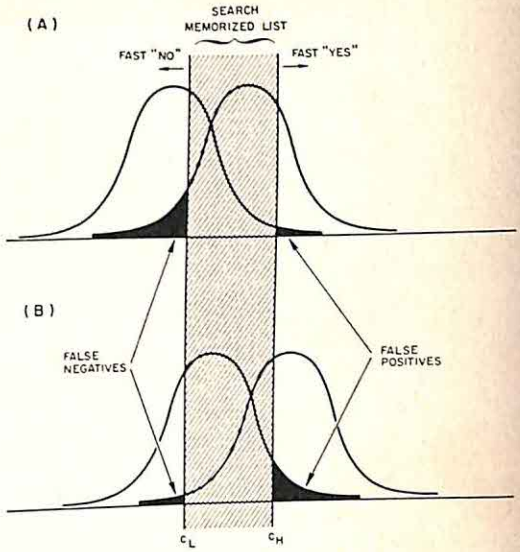

FIG. 2. Distributions of subjective familiarity values for distractor words (left) and target words (right), on the familiarity continuum. (Figure 2A represents the relative locations of the distributions at the start of the session, whereas Fig. 2B shows the increase in the means of both distributions, after the presentation of a specific test word.)

There are several possible procedures that S could use in searching the target list. A self-terminating, serial search of the entire list might seem unlikely, since in both the present study and in Juola et al. (1971) the serial position curves were essentially flat. This lack of serial position effects could reflect either an exhaustive scan of the list or a self-terminating scan with a random starting point. If whenever familiarity fell between c_L and c_H, S scanned the entire target list before responding, the time taken for the search would be equal for targets at the beginning and end of the list. Alternatively, S may not begin his search at the start of the list, but at a random point, retrieving a portion of the list and scanning it exhaustively; then this process is repeated if the test item is not found. Here again, targets at the beginning of the list would, on the average, have no advantage over those at the end.

In each block, test repetitions have a smaller effect on negative latency than on positive latency. One explanation for this effect is that in accessing the stored information about a test word, S also sometimes retrieves the response that was made to that word on an earlier pres-

entation. If response memory can be used to make a rapid decision to output the same response made previously, then the effect of repeated tests should be less for distractor words than for targets. This is true because both response memory and an increased familiarity rating for repeated target words would decrease the probability that *S* has to scan through the memorized list before responding. For repeated distractors the higher familiarity value increases the probability of a slower response following the outcome of the list search; however, this effect is attenuated if *S* can sometimes make a more rapid decision based on his response memory.

The present error data are similar to those obtained by Underwood and Freund (1970) in a forced-choice recognition study. There, repetition of distractors produced more errors, while repeating targets produced fewer mistakes. As in the present study, then, increasing an item's experimental frequency made it more likely that it would be called a target item.

Thus, while familiarity (or frequency) is certainly not the only information available to *S* about a stimulus, it is apparently an attribute of some importance and generality in the recognition task. It is when familiarity is made a doubtful cue that *S* must access further information in order to respond accurately.

REFERENCES

Atkinson, R. C., & Juola, J. F. Factors influencing speed and accuracy of word recognition. (Tech. Rep. No. 177) Palo Alto, Calif.: Stanford University, Institute for Mathematical Studies in the Social Sciences, 1971.

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*. Vol. 2. 1968.

Bertelson, P. S–R relationships and reaction times to new versus repeated signals in a serial task. *Journal of Experimental Psychology*, 1963, **65**, 478–484.

Hintzman, D. Recognition time: Effects of recency, frequency, and the spacing of repetitions. *Journal of Experimental Psychology*, 1969, **79**, 192–194.

Juola, J. F., Fischler, I. S., Wood, C. T., & Atkinson, R. C. Recognition time for information stored in long-term memory. *Perception & Psychophysics*, 1971 (in press).

Kintsch, W. Memory and decision aspects of recognition learning. *Psychological Review*, 1967, **74**, 496–504.

Parks, T. E. Signal detectability theory of recognition memory performance. *Psychological Review*, 1966, **73**, 44–58.

Shepard, R. N., & Teghtsoonian, M. Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology*, 1961, **62**, 302–309.

Sternberg, S. High-speed scanning in human memory. *Science*, 1966, **153**, 652–654.

Underwood, B. J., & Freund, J. S. Testing effects in the recognition of words. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 117–125.

(Received March 20, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 59–65

SEMANTIC GRADIENTS AND INTERFERENCE IN NAMING COLOR, SPATIAL DIRECTION, AND NUMEROSITY[1]

LESLIE A. FOX, RONALD E. SHOR, AND ROBERT J. STEINMAN[2]

University of New Hampshire

Using generalizations of the Stroop test, interference in naming pictorially represented color, spatial direction, and numerosity was measured for stimulus charts differing in the semantic relation of irrelevant nonpictorial stimulus aspects to the relevant pictorial aspects in a repeated-measures design with 112 *S*s. Amounts of interference exhibited with the color charts followed a gradient corresponding to strength of semantic relationship, all differences being significant, and correlated very highly with results obtained by G. S. Klein in 1964. For numerosity charts, all differences in interference were significant and were in directions according with a semantic gradient. For spatial direction charts, directions of all significant differences corresponded to a semantic gradient, but some nonsignificant differences were not in expected directions. Particular comparisons of interference exhibited on certain numerosity charts further support hypotheses that the interference is a function primarily of semantic relationship rather than of frequency of usage and that it occurs in response production rather than in stimulus reception.

One of the most reliable psychological phenomena is that it is more difficult to name the colors of inks when the inks are used to print the names of different colors than when the inks are used in connection with nonmeaningful material. For example, naming the color of red ink used to print the word BLUE is more difficult than naming the color of a red patch. This color-word interference effect was first studied by Stroop (1935).

Analogous effects of interference between pictorial and nonpictorial representations of concepts have been found to occur in concept domains other than color. Windes (1968) found that the time taken to name the numerosities of a series of groups of one, two, or three symbols was significantly longer when the symbols were the numerals 1, 2, or 3 than when the symbols were plus signs. Morton (1969) found essentially the same interference effect when *S*s were required to sort cards into bins labeled with Arabic numerals according to numerosities of stimuli on the cards. Morton used the numerosities one through six and used the number names ONE through SIX as well as the Arabic numerals 1 through 6 as incongruent stimulus aspects. In addition, Morton discovered that an analogous interference effect occurs in the domain of spatial position in that the time taken to sort a set of cards according to stimulus positions was significantly longer when the stimuli were the position names RIGHT, LEFT, and CENTER appearing in incongruent positions than when the stimuli were groups of plus signs. Shor (1970, 1971) has found that interference occurs in a large number of concept domains, including naming the numerosities one through six when stimuli are presented in standard die-face patterns and naming the spatial directions up, down, right, and left when the pictorial representation consists of "minimal" arrows (rectangles with single pointed ends) and the conflicting nonpictorial representation consists of the direction names UP, DOWN, RIGHT, and LEFT.

Klein (1964) conducted a study which has shed additional light on the nature of the interference found in the Stroop-type color-naming task. The results of Klein's study suggest that the degree of interference caused by irrelevant nonpictorial information follows a gradient that is a

[1] This study was supported in part by Grant MH 16209-01 from the National Institute of Mental Health.

[2] The authors wish to thank G. Alfred Forsyth and George M. Haslerud for their helpful comments and Richard A. Bagg and Richard I. Kushner for their assistance in data collection.

function of the strength of the semantic relation between the irrelevant information and the names of the colors. Klein used seven lists of stimuli which, by intuitive considerations, differed in strength of this semantic relationship. The nonpictorial stimulus aspects of the seven lists were: (*a*) names of the colors appearing pictorially (RED, GREEN, YELLOW, BLUE) but printed in incongruent colors; (*b*) names of colors not appearing pictorially (TAN, PURPLE, GREY, BLACK); (*c*) color-associated words (LEMON, GRASS, FIRE, SKY) printed in incongruent colors; (*d*) noncolor-associated common words (PUT, HEART, TAKE, FRIEND); (*e*) rare words (SOL, HELOT, EFT, ABJURE); (*f*) nonsense syllables (HJH, EVGJC, BHDR, GSXRQ); and (*g*) groups of asterisks. In a between-*S*s design, Klein found that interference caused by the lists, using total response time as the basic dependent measure, decreased in the order given above with differences between adjacent Lists *a* and *b*, *d* and *e*, and *f* and *g* being significant ($p < .05$).

The semantic gradient interpretation of interference in a Stroop-type task, i.e., that degree of interference is a positive function of semantic relatedness, received further support in subsequent studies. Bakan and Alperson (1967) showed that the interference is not a function of pronounceability of the irrelevant nonpictorial aspects of the stimuli. Bakan and Alperson used five sets of four trigrams each, the sets differing in pronounceability. Of the five sets, one consisted of nonsense trigrams matched for mean pronounceability with another set consisting of three-letter words; the word set caused significantly more interference than did the nonsense trigram set. Scheibe, Shaver, and Carrier (1967) used an objective method for assessing semantic relatedness prior to obtaining color-naming data. They used five sets of five words each, the irrelevant nonpictorial stimulus aspects for the five sets being: (*a*) names of the common colors appearing (PURPLE, BROWN, BLUE, GREEN, RED) but printed in incongruent colors; (*b*) associated uncommon color names (VIOLET, TAN, NAVY, EMERALD, SCARLET) printed in incongruent associated colors; (*c*) names of uniformly colored objects or substances (GRAPE, CHOCOLATE, SKY GRASS, BLOOD) printed in incongruent colors; (*d*) names of ambiguously colored objects or substances (PAINT, PENCIL, CARPET, COLOR, CAR); and (*e*) names of abstractions or nonreferential parts of speech (REFER, IDEAL, AND, BELIEVE, REASON). The latter four sets, printed in black ink, were presented in a word-association task, and for each set the relative frequency of the color names of Set *a* as associates was taken as the degree of semantic relatedness of that set. Sets *b* and *c* did not differ appreciably on this measure, and in a subsequent color-naming task with different *S*s, these sets did not differ significantly in amount of interference; similarly, Sets *d* and *e* did not differ appreciably on the color-association measure and did not differ significantly in amount of interference with color naming.

Since the basic interference effect occurs in the domains of numerosity and spatial position and direction as well as of color, it is natural to ask whether the semantic gradient effect found in the domain of color occurs with numerosity and spatial position and direction also. Morton (1969) obtained some evidence that indicates that the semantic gradient effect occurs in the domain of numerosity. In a number of independent studies involving card sorting according to numerosity of stimuli, Morton obtained results which, taken together, suggest that (*a*) Arabic numerals corresponding to the numerosities used but appearing in incongruent numerosities, (*b*) noncorresponding Arabic numerals, (*c*) letters, and (*d*) nonsense symbols cause degrees of interference that decrease in the order listed. In another study, Morton found that common words (WHITE, LOW, EASY, FULL, HIS, BIG) caused more interference than did corresponding number names (ONE through SIX), but it should be noted that the task involved sorting cards into bins labeled with Arabic numerals.

The present study is an attempt to determine whether the semantic gradient of interference occurs in the concept do-

mains of spatial direction and numerosity, using stimuli designed to be as nearly analogous as possible to the stimuli used by Klein (1964) and to confirm the existence of the effect in the domain of color. Although Morton (1969) has provided evidence relevant to numerosity, it involves a sorting task rather than a naming task and is somewhat fragmented in nature. The present study used a naming task and presented Roman numerals as nonpictorial representations of numerosity in addition to number names and Arabic numerals.

METHOD

Subjects

The *S*s were 112 students enrolled in introductory psychology classes at the University of New Hampshire, in which participation as *S*s in psychological experiments was required of all students.

Apparatus

All stimulus charts were constructed on 71.12 $\times$ 55.88 cm. white posterboard. Each chart had 80 stimuli arranged in 10 rows of 8 stimuli each with columns and rows on 6.35-cm. centers. All letters and numerals were printed with rubber stamps having a 3.175-mm. sanserif typeface; all letters were capitals. All charts were displayed on a rack inclined 20° from vertical, placed on a table of standard height. A Hunter Model 120A timer was used in timing *S*s' performance.

Color charts.—On all charts used in the color portion of the experiment, stimuli consisted of words or groups of letters printed in black, blue, green, and red ink, with 20 stimuli in each color on each chart. The words or letter groups appearing on the seven charts were as follows: (A) same color names (BLACK, BLUE, GREEN, RED) printed in incongruent colors, (B) different color names (GRAY, WHITE, BROWN, YELLOW), (C) color-associated nouns (COAL, SKY, GRASS, BLOOD), (D) common words (HOUSE, FOOT, CHAIR, HAT), (E) rare words (MIDGE, NEWT, FLUME, CUD), (F) consonant groups (TLRLK, TKKT, JRKBM, HVL), and (G) undifferentiated groups of letter Os (OOOO). Ink colors and words or letter groups were randomly ordered except for certain restrictions. Stimuli were organized in 20 blocks of 4 stimuli each in standard reading order with each block constituting a random permutation of the colors and of the words or letter groups (except on Chart G) with the restriction that no color and no word or letter group appear twice in succession and, on Chart A, with the restriction of incongruity. All words or letter groups on Charts D, E, and F were pair-wise matched with words on Chart A for number of letters. All words on Charts A, B, C, and D were classified AA in frequency of usage by Thorndike and Lorge's *The Teachers Word Book of 30,000 Words*, and all words on Chart E were classified as occurring at least once and less than twice per million words by Thorndike and Lorge. The groups of consonants on Chart F were chosen randomly.

Direction charts.—Stimuli on the charts used in the direction portion of the experiment were 5.08 $\times$ 5.08 cm. squares drawn in black ink, each containing a word or group of letters printed in black ink near the top, bottom, right, or left of the square, with 20 squares containing words or letter groups in each of the four positions. The words or letter groups appearing on the seven charts were the following: (A) same direction names (UP, DOWN, RIGHT, LEFT) appearing in incongruent positions, (B) different direction names (NORTH, SOUTH, EAST, WEST), (C) direction-associated verbs (LIFT, DROP, TURN, FLOW), (D) common words (IT, GRAY, HEART, BOOK), (E) rare words (PI, DIRK, STOAT, TARN), (F) consonant groups (JB, FWXD, TRTDR, RDLN), and (G) undifferentiated groups of four letter Os each. Positions within the squares and words or letter groups were randomly ordered in a manner similar to that for colors and words or letter groups on the color charts. All words or letter groups on Charts D, E, and F were pair-wise matched with words on Chart A for number of letters. All words on Charts A, B, C, and D were classified AA in frequency of usage by Thorndike and Lorge. DIRK, STOAT, and TARN occur at least once and less than twice per million words, while PI occurs six times per 18 million words, according to Thorndike and Lorge. The groups of consonants on Chart F were chosen randomly.

Numerosity charts.—Stimuli on the charts used in the numerosity portion of the experiment were 5.08 $\times$ 5.08 cm. squares drawn in black ink, each containing one to six identical symbols or words arranged in a standard die-face pattern. The symbols or words used on the eight charts were as follows: (A) same number names (ONE through SIX) appearing in incongruent numerosities, (B) same Arabic numerals (1 through 6) appearing in incongruent numerosities, (C) same Roman numerals (I through VI) appearing in incongruent numerosities, (D) different Arabic numerals (7 through 12), (E) common words (AND, BED, GRASS, TAKE, BLUE, CRY), (F) letters (A through F), (G) abstract symbols, and (H) circles. All of the symbols on Chart G and the circles on Chart H were 4 mm. high and were drawn in black ink using templates; the symbols used on Chart G are shown in Fig. 1. On all charts, stimuli were organized in 10 blocks of six stimuli each and 1 block of two stimuli in standard reading order. Each block of six constituted a complete random permutation of the six

FIG. 1. Abstract symbols used on Numerosity Chart G.

numerosities and of the symbols or words (except on Chart H) with the restriction that no numerosity and no symbol or word appear twice in succession and, on Charts A, B, and C, with the restriction of incongruity. The last block of two stimuli constituted a random sample without replacement. Words on Chart E were pair-wise matched with words on Chart A for number of letters. All words on Charts A and E were classified AA in frequency of usage by Thorndike and Lorge.

Design and Procedure

Order of presentation of the three chart sets was completely counterbalanced. Presentation order of the charts in each set was determined by means of balanced Latin squares.

Each *S* was seated initially approximately 90 cm. from the rack holding the charts. A preliminary vision test used a chart made of 71.12 × 55.88 cm. white posterboard on which appeared four words, printed in black ink, and four groups of letter os printed in the four colored inks used on the experimental charts; all letters were of the same typeface that was used on the experimental charts. The *S* was instructed to read aloud the four words and to name the four ink colors; all *S*s were able to comply. The *S* was then allowed to adjust the position of the chair if he desired.

Before each chart set was presented, *S* was instructed in the task for that set and was shown a 35.56 × 27.94 cm. white posterboard chart on which appeared a sample stimulus from each of the charts in the set. The *S* was instructed to complete each chart as rapidly as possible, trying not to make errors and not to correct errors. For the color charts, *S* was instructed to ignore the words or letter groups as such and to name the ink color of each stimulus in standard reading order. For the direction charts, *S* was instructed to ignore the words or letter groups as such and to name the position of the word or letter group in each square in standard reading order using the words UP, DOWN, RIGHT, and LEFT. For the numerosity charts, *S* was instructed to ignore the symbols or words as such and to name the number of symbols or words in each square in standard reading order.

The *S* turned his head away from the charts while *E* removed the last chart used and, after *E* announced "ready," was allowed as much time as he desired to prepare himself before turning to face forward and begin the task for the next chart. For each chart, *E* recorded the time taken and errors committed for Rows 2–10.

RESULTS

The means and standard deviations for all of the charts used are shown in Table 1 for time data and in Table 2 for error data. As might be expected with elasped time and frequency data, there are strong positive relationships between the means and the variances, indicating heterogeneity of variance. Following accepted practice (Winer, 1962), the time data were subjected to a logarithmic transformation, and the error data were subjected to a square-root transformation. In analyses of the transformed data, it was found that frequency of errors was a less sensitive measure than elapsed time was, for there were many fewer significant differences between error means than there were between time means. Wherever significant differences between error means occurred, there were corresponding significant differences between time means, and wherever differences between error means and differences between time means were in opposite directions, the differences between error means were not significant ($p > .05$). Therefore, the error data were not taken into consideration in the conclusions reached.

Analysis of variance of the time data after the logarithmic transformation showed the

TABLE 1

CHART MEANS AND STANDARD DEVIATIONS FOR TIME DATA

Chart	Color		Direction		Numerosity	
	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
A	77.273	11.559	60.507	10.707	51.399	9.306
B	69.186	11.740	54.471	10.097	49.690	7.923
C	60.145	10.726	53.017	9.713	46.521	8.651
D	55.630	8.860	53.401	9.971	43.985	7.608
E	52.441	8.129	51.143	9.212	42.395	7.222
F	50.184	7.091	51.360	9.009	41.197	7.075
G	47.804	7.490	50.024	9.115	36.848	6.754
H					35.427	6.149

Note.—Data recorded in seconds.

TABLE 2

CHART MEANS AND STANDARD DEVIATIONS FOR ERROR DATA (FREQUENCIES)

Chart	Color		Direction		Numerosity	
	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
A	3.64	3.89	1.94	2.15	1.66	1.99
B	2.20	2.56	1.54	1.77	1.53	1.55
C	1.38	1.67	1.49	1.98	.85	1.14
D	1.10	1.49	1.35	1.75	.67	1.12
E	1.11	1.59	1.10	1.57	.42	.79
F	.90	1.11	1.27	1.59	.40	.75
G	1.26	1.25	1.62	2.26	.24	.52
H					.16	.44

differential effects of charts to be significant for the color charts, F (6, 666) = 633.43, the direction charts, F (6, 666) = 148.21, and the numerosity charts, F (7, 777) = 406.93, all ps $< .001$. The Newman-Keuls method (Winer, 1962) was used in a posteriori tests of mean differences of log times. The order of mean times of the color charts corresponds with the semantic gradient (i.e., A through G in decreasing order); all differences between means are significant ($p < .01$). The order of mean times of direction charts, largest first, is A, B, D, C, F, E, and G. All differences between nonadjacent means are significant ($p < .01$). Significance levels for differences between adjacent means are as follows: A versus B, $p < .01$; B versus D, $p < .05$; D versus C, $p > .05$; C versus F, $p < .01$; F versus E, $p > .05$; E versus G, $p < .05$. The order of mean times of numerosity charts is A through H in decreasing order; all differences between means are significant ($p < .01$).

Discussion

The effects of the color charts were clearly in accord with those found by Klein (1964) and were consistent with the semantic gradient interpretation of interference effects. Moreover, the chance probabilities of the present results are lower than those of Klein's, probably as a result of substantially larger sample sizes and the inherently greater power of the repeated-measures design.

Since the color portion of the experiment was designed to be a relatively straight-forward replication of Klein's (1964) study, a comparison of results clearly has bearing on the reliability of the phenomenon under examination. However, important differences in procedure and stimuli should first be noted. First, Klein timed performance on 80 stimuli, while in the present study performance was timed on only the last 72 stimuli; hence, adjustment of the mean times reported here by multiplying by 80/72 is necessary for comparability. Second, all *S*s in Klein's study were presented with only three pages of stimuli, always in the order: warm-up (consisting of color names printed in black), colors-alone, and experimental. Thus, his colors-alone page (corresponding to Chart G) was never preceded by an experimental page, and no stimulus page was preceded by more than two other pages. By contrast, in the present study each color chart was presented in balanced order with 21 other charts, so that each color chart was preceded by 10.5 other charts on the average. On the basis of this second consideration, one would predict some general systematic difference in results due to either practice or fatigue effects, especially between Color Chart G and Klein's colors-alone page, although on this pair the effect is confounded with another factor explained below. It follows from these first two points that evidence regarding the reliability of the phenomenon and the nature of order effects would be brought out more clearly by regression of Klein's means on the means from the present study adjusted as mentioned above. Regression on the adjusted means would yield evidence bearing exclusively on reliability and order effects, for adjustment would compensate mathematically for the clearly specifiable effects of the difference in number of stimuli over which performance was timed. The resulting additive constant of the regression analysis would indicate only the nature and magnitude of order effects, for it would not be influenced by the difference in number of stimuli. Degree of proximity to unity of the resulting regression coefficient would indicate the reliability of the phenomenon independent of effects of systematic differences in procedure. Third, two of the four words on Klein's distant color-words page, PURPLE and TAN, occur less frequently in normal usage than do the words on corresponding Color Chart B according to Thorndike and Lorge. The mean times for Chart D (common words) and Chart E (rare words) demonstrate that when color relatedness is held constant, interference with color naming is a positive function of frequency of usage. It follows that the mean for Klein's distant color-names page should be appreciably less than the linear regression estimate based on the adjusted mean for Chart B. Fourth, stimuli on Klein's colors-alone page consisted of rows of asterisks, while stimuli on corresponding Chart G consisted of groups of four os. Although one might predict greater interference on Chart G than on the colors-alone page because a group of four os seems to have greater semantic content than a row of asterisks does, this factor is confounded with a possible practice or fatigue effect mentioned in the second point above.

The mean times from Klein's (1964) study are shown in the second column of Table 3, and the adjusted mean times for the corresponding

TABLE 3

COMPARISON OF KLEIN'S (1964) MEAN TIMES WITH ADJUSTED MEAN TIMES OF CORRESPONDING COLOR CHARTS AND LINEAR REGRESSION ESTIMATES

Klein (1964)		Estimates[a]		Corresponding chart	
Stimulus page	$\bar{X}$	Unselected basis[b]	Selected basis[c]	Letter	Adjusted $\bar{X}$[d]
Color-names close	81.47	77.54	81.07	A	85.86
Color-names distant	62.17	68.79	71.48	B	76.88
Color-related meanings	59.43	58.99	60.73	C	66.82
Common words	55.95	54.10	55.38	D	61.81
Rare words	51.18	50.65	51.60	E	58.27
Nonsense syllables	49.68	48.20	48.92	F	55.76
Colors alone	44.02[e]	45.62	46.09	G	53.11

Note.—Mean times and estimates are recorded in seconds.
[a] Estimates from linear regression of Klein's means on adjusted means for corresponding color charts.
[b] Regression equation derived from all seven pairs of means.
[c] Regression equation derived from all pairs of means except those for Charts B and G.
[d] Mean for corresponding chart adjusted for comparability by multiplying by 80/72.
[e] Averaged over six independent groups of 15 *S*s each.

charts of the present study are shown in the last column. For all seven pairs of means, the correlation is .961, the regression coefficient is .975, and the additive constant is -6.147; the resulting regression estimates of Klein's means are shown in the third column of Table 3. By the arguments above, the pair for Klein's distant color-names page and Chart B and the pair for Klein's colors-alone page and Chart G are less comparable than the remaining five pairs are, so the analysis was also performed for only these five pairs of means, yielding a correlation of .998, a regression coefficient of 1.068, and an additive constant of -10.633; the resulting estimates of Klein's means are shown in the fourth column of Table 3. By either set of values, it is clear that the phenomenon is very reliable. The values of the regression coefficient are very close to unity, indicating an impressively high consistency of effects on different groups of *S*s. The values of the additive constant indicate that there was an appreciable general fatigue effect in the present study with the color charts and, presumably, with the direction and numerosity charts. The mean for Klein's distant color-names page is substantially lower than either of the regression estimates based on the adjusted mean for Color Chart B, substantiating the argument that the distant color-names page caused less interference than Chart B did because of the relatively low frequency of usage of PURPLE and TAN.

The effects of the direction charts were not completely in accord with predictions based on the semantic gradient hypothesis, but the contradictions that occurred were not clear-cut. The order of mean times of the charts corresponded with the order that would be expected according to the semantic gradient hypothesis except for the reversal of Charts C (direction-associated verbs) and D (common words) and the reversal of Charts E (rare words) and F (consonant groups). Differences between adjacent means that did follow the expected order were significant ($p < .05$), whereas the differences between means for Charts C and D and for E and F were not significant ($p > .05$). The nonsignificant reversal of the means for Charts C and D might be the result of an inherent ambiguity or symmetry on the lateral direction axis. Gravity sensors provide an inherent basis for distinguishing between up and down, but nothing analogous exists for distinguishing between right and left; indeed, some *S*s volunteered remarks about their weak ability to distinguish right from left. This symmetry on the lateral axis is reflected in the lack of verbs that are unambigously associated with either right or left in the same way that COAL, SKY, GRASS, and BLOOD are associated with the colors black, blue, green, and red, respectively. Hence, the average semantic distance of the words on Direction Chart C from the words on Direction Chart A may be considerably greater than that of the words on Color Chart C from the words on Color Chart A. As a result, assuming the semantic gradient hypothesis to be true for directions as well as for colors, Direction Charts C and D would be more alike in magnitude of interference than would be the case for Color Charts C and D. If this were true, a reversal of the means for Direction Charts C and D because of sampling error

would be less unlikely. It follows from this line of argument, also, that results more clearly in accord with the semantic gradient hypothesis would result if the experiment were repeated with directions restricted to up and down, with different-direction names consisting of RIGHT and LEFT, with direction-associated verbs restricted to LIFT and DROP, and so on.

The effects of the numerosity charts were entirely in accord with the semantic gradient hypothesis, and three points should be considered in particular. First, Chart A (same number names) caused significantly more interference than did Chart E (common words), which directly contradicts part of Morton's (1969) results despite the difference in tasks but is based on a substantially greater sample size. Second, a comparison of the amounts of interference caused by Charts D (different Arabic numerals) and C (same Roman numerals) is especially important, for an explanation based solely on frequency of usage would predict that D would cause more interference than C would, while an explanation taking semantic content as the primary factor would predict the opposite; in fact, the results followed the latter prediction. Third, a comparison of the amounts of interference caused by Charts B (same Arabic numerals) and A has implications regarding the locus of the interference. The fact that conflicting nonpictorial representations of a linguistic nature (Chart A) caused more interference than did those of a nonlinguistic nature (Chart B) lends support to theories holding that the interference occurs in the production of the response rather than in the reception of the stimulus. In terms of the numerical information conveyed, the visual symbols 1 and ONE, for example, are equivalent, but ONE is more closely associated with the spoken response "one," and with all other spoken number names, than 1 is. Hence, when *S* is attempting to name numerosities, two 1's contain the same amount of numerical information as two ONEs do, but the nonpictorial information of the two ONEs will interfere more than that of two 1's in the process of translating the pictorial information into the correct spoken response "two."

In summary, the experiment showed, first, that the interference gradient found by Klein (1964) in the concept domain of color has impressively high reliability and, second, that analogous interference gradients occur in the domains of spatial direction and numerosity. Furthermore, results in the numerosity portion of the experiment provided strong evidence in support of the semantic gradient hypothesis that the interference is a function of semantic relatedness. Comparison of the degrees of interference caused by different Arabic numerals and by same Roman numerals strongly indicates that the interference is a function primarily of the semantic content of the nonpictorial information rather than of frequency of usage. In addition, comparison of amounts of interference caused by same Arabic numerals and by same number names suggests that the interference is also a function of the relationship of the nonpictorial stimulus aspect code to the response code and, hence, that it occurs in the production of the response rather than in the reception of the stimulus. Results of the direction portion of the experiment were less clear-cut, although all significant differences were in accord with semantic gradient expectations. One of the contradictory, nonsignificant differences may have been the result of the lack of a fundamental distinction between right and left analogous to that between up and down.

REFERENCES

BAKAN, P., & ALPERSON, B. Pronounceability, attensity, and interference in the color-word test. *American Journal of Psychology*, 1967, **80**, 416–420.

KLEIN, G. S. Semantic power measured through the interference of words with color-naming. *American Journal of Psychology*, 1964, **77**, 576–588.

MORTON, J. Categories of interference: Verbal mediation and conflict in card sorting. *British Journal of Psychology*, 1969, **60**, 329–346.

SCHEIBE, K. E., SHAVER, P. R., & CARRIER, S. C. Color association values and response interference on variants of the Stroop test. *Acta Psychologica*, 1967, **26**, 286–295.

SHOR, R. E. The processing of conceptual information on spatial directions from pictorial and linguistic symbols. *Acta Psychologica*, 1970, **32**, 346–365.

SHOR, R. E. Symbol processing speed differences and symbol interference effects in a variety of concept domains. *Journal of General Psychology*, 1971, in press.

STROOP, J. R. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 1935, **18**, 643–662.

WINDES, J. D. Reaction time for numerical coding and naming of numerals. *Journal of Experimental Psychology*, 1968, **78**, 318–322.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received March 20, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 66–73

EFFECTS OF PRIOR FREE RECALL TESTING ON FINAL RECALL AND RECOGNITION [1]

CHARLES F. DARLEY [2]
Stanford University

AND

BENNET B. MURDOCK, JR.
University of Toronto

The *S*s were shown 10 lists of words, and each list was either followed by an immediate free recall test or was not tested. The *S*s then received either a final free recall test on the words from all 10 lists or a three-alternative forced-choice recognition test. Initial testing facilitated retrieval on the final recall test for all serial positions but had no overall effect on recognition performance. Nonpositive recency occurred on final testing of items from both tested and nontested lists whether the test was recall or recognition. A number of words presented on lists prior to immediately tested lists were incorrectly included in immediate recall for the tested lists. These prior list intrusions were largely items which had been presented in lists which received no immediate test. The frequency of these prior list intrusion errors decreased over their original serial input positions. Final recall performance was an increasing function of output position in initial recall, although there was an indication that items from final output positions were recalled less well on the final test than those from intermediate output positions. The main conclusion is that prior testing increases item accessibility but not availability.

The phenomenon of negative recency was uncovered in a free recall memory study performed by Craik (1970). Craik presented *S*s with ten 15-word lists with an immediate free recall test following each list. Following recall of the tenth list, *S*s were instructed to recall as many words as possible from all 10 lists. Craik found that the positive recency effect, a common feature of serial position curves for immediate recall, disappeared from the final recall test. In fact, final recall performance for items from the last serial positions was poorer than for other items.

The present study seeks to clarify the nature of the negative recency effect. Previous studies examining performance on a final test following presentation of a series of word lists employed immediate recall tests of the lists. It remains a question whether results similar to Craik's (1970) would be obtained if word lists received no immediate test. Craik, Gardiner, and Watkins (1970) conditionalized on immediate recall or nonrecall of items when examining final recognition performance, but because of possible selection artifacts, it is more desirable to manipulate the tested versus nontested factor explicitly. This manipulation was performed by Madigan and McCabe (1971), but a paired-associate (PA) probe paradigm was used. They obtained a negative recency effect on a final recall test for pairs which received no immediate test as well as for those which did receive a test.

In addition, there are indications that performance on a final recognition test differs markedly from performance on a final recall test. Cohen (1970) presented a final recall test followed by a final recognition test to *S*s who had seen a series of immediately tested word lists. Cohen showed that the curve for final recognition performance as a function of serial position of initial presentation differs in shape from the serial input position curve for performance on a final recall test. However, the positive recency effect which Cohen obtained for final recognition testing was not confirmed by Rundus, Loftus, and Atkinson (1970). Rundus et al. presented a

[1] This research was supported by United States Public Service Grant 1-F01-MH-40, 397-01 PS from the National Institute of Mental Health. The authors thank Richard C. Atkinson, Dewey Rundus, and Thomas O. Nelson for advice concerning the design of this study and the interpretation of the results, and Kathy Troyer for help in analyzing the data.

[2] Requests for reprints should be sent to Charles F. Darley, Department of Psychology, Stanford University, Stanford, California 94305.

3-wk. delayed recognition test on words Ss had seen in a series of lists which had received immediate free recall tests. Rundus et al. found that the serial position curve for the probability of delayed recognition showed a negative recency effect. McCabe and Madigan (1971) also obtained a negative recency effect for both stimulus and response recognition on a final test of PA pairs.

Combining in a single free recall study the immediate recall variable with the type of final test should clarify whether there are differences between final recall and final recognition tests with regard to negative recency. If there are such differences, such a study would show whether the differences are present for words from initially nontested lists as well as for words from initially tested lists. Also, any possibilities that one type of final test will affect performance on the other, a factor which could be present in Cohen's (1970) study, would be eliminated.

In the present study, Ss studied a series of ten 20-word lists, half of which were followed by an immediate free recall test and half by a period of counting forward by threes. After the tenth list had been presented and either an immediate test on that list had been administered or a period of counting following the list presentation had been concluded, half of the Ss received a final free recall test and half received a final recognition test.

METHOD

Forty-eight Stanford University students were each paid $2.00 to serve as Ss. Half were females and half were males. Each S was shown 11 different lists of 20 unrelated nouns with frequencies of occurrence from 10 to 40 per million (from *The Teacher's Word Book of 30,000 Words* by Thorndike & Lorge). The first list was a practice list with the same words in the same order for all Ss. The order of the following 10 lists and the order of items within each list were random for each S. Words were presented singly on cards, each word being shown for 5 sec. The S's rehearsal process for each list was recorded using the procedure described by Rundus and Atkinson (1970). The S was instructed to study the list by spending the 5-sec. presentation interval rehearsing aloud any words he had seen in the present list, including the currently presented word. This overt rehearsal was tape recorded. The overt rehearsal procedure was used in order to make the present study as comparable as possible to that of Rundus et al. (1970). The overt rehearsal data were not analyzed and will receive no further consideration in this paper.

Immediately after presentation of each list, S was either tested on the list or counted forward by threes for a short period of time. The S did not know whether a list would receive a test or not until after the last word of the list had been studied. The test was a written free recall test for which S received as much time as he needed to write the recalled words. The Ss usually completed the free recall tests in approximately 2 min. For untested lists, Ss counted for a length of time equal to the time they had taken to complete the free recall test on the most recent tested list. If S had received no free recall tests on earlier lists, he counted for 2 min. Whether a list would be tested or untested was determined randomly with the constraints that (*a*) no more than three tested or nontested lists could occur consecutively and (*b*) of the 10 lists, 5 were tested and 5 were not.

Subsequent to the test or the period of counting following the last list, Ss counted backward by twos for 1 min., starting with a number supplied by E. Half of the Ss were then given as much time as they required to recall as many items as they could. Twelve of the Ss receiving this final recall test were males and 12 were females. The other 24 Ss, half males and half females, were given a 200-item three-alternative forced-choice recognition test on the words from the 10 lists. Ten test sheets were given to S, each sheet containing 20 rows of three words per row. One word from each row had been presented on one of the lists, while the other two words were lures drawn from the same population as the previously presented words. Each S received the same order of words on the recognition test. The S was instructed to circle the one word in each row which he had seen earlier and to give a confidence rating using the numbers 1, 2, or 3, where 3 was highly confident and 1 was a guess. Neither the Ss who received a final recall test nor those who received a final recognition test were informed that they would receive a final test on the words.

RESULTS AND DISCUSSION

Serial position curves for immediate recall, final recall, and final recognition are shown in Fig. 1. The immediate recall curve represents data both from Ss who received a final free recall test and from those who were given a 200-item three-alternative forced-choice recognition test. The data are pooled because recall Ss and recognition Ss were virtually identical in both the overall level of performance

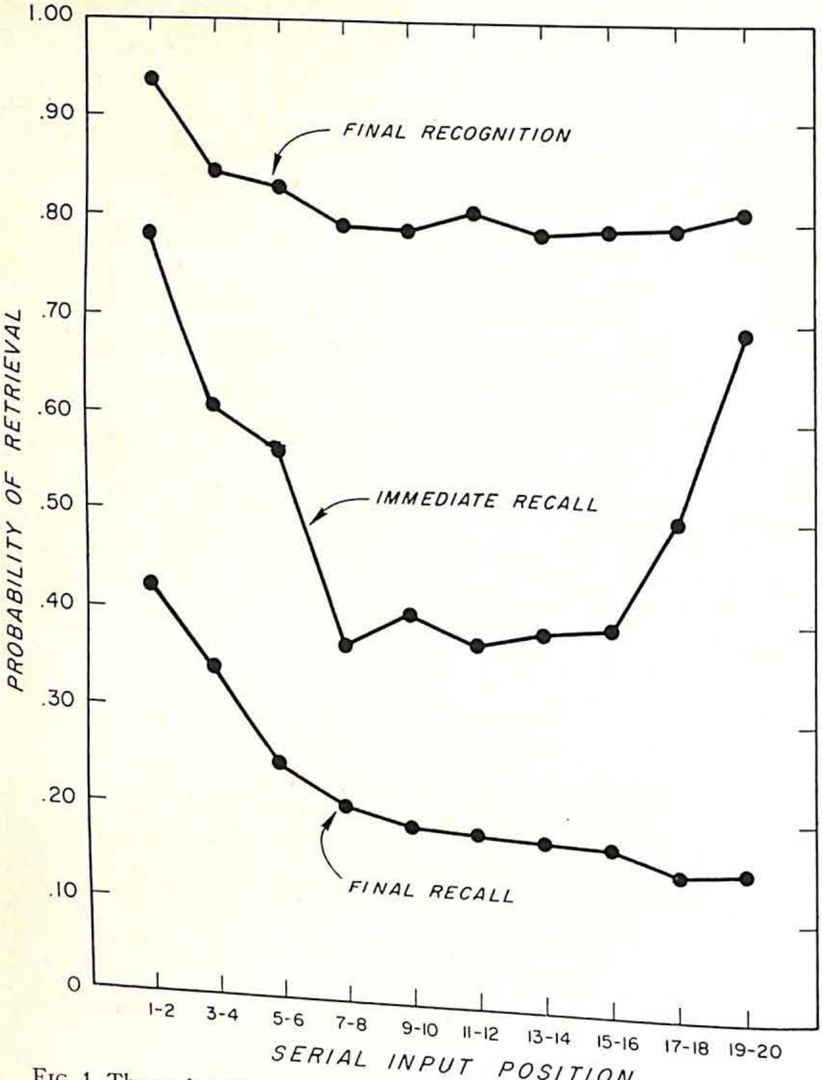

FIG. 1. The probability of initial recall, final recall, and final recognition as functions of an item's position during study.

(50.7% vs. 48.5%, respectively) and the shape of the serial position curves. The immediate recall curve shows the typical primacy and recency effects with depressed performance for items in the middle of the list. Recency is not as pronounced as in most studies, probably due to both the relatively slow presentation rate and the overt rehearsal procedure (cf. Rundus & Atkinson, 1970).

Data for the final recall and recognition curves in Fig. 1 are from all items presented to *S*s. In other words, all items from each of the 10 study lists are included in these curves, with no distinction made between items from initially tested and initially nontested lists. The final recall curve shows pronounced primacy, with performance decreasing over serial positions. The existence of a negative recency effect was confirmed by a linear trend analysis for Positions 14–20. The decreasing linear trend was significant, F (1, 138) = 4.33, $p < .05$. Final recognition is similar, but following the primacy effect performance was relatively constant rather than decreasing or increasing over serial positions, as shown by the nonsignificant linear trend, F (1, 138) = .69, for Serial

Positions 14–20. Note that the chance level of performance is .33.

Figure 2 presents serial position curves for final recall and final recognition with separate curves for items from initially tested and initially nontested lists. Final recall for items from both tested and nontested lists shows primacy and decreasing performance over serial positions. However, due to the reduced number of observations, this decrease over the terminal positions was not significant. For tested items, F (1, 138) = 1.23, and for nontested items, F (1, 138) = 3.19. Overall performance was significantly higher for tested items (27.5%) than for nontested items (17.6%), t (23) = 4.43, $p < .01$. This indicates that receiving an initial recall test facilitates later recall of an item regardless of the item's original input position.

One can ask whether an item from the terminal input positions receives the same facilitation in final recall due to initial testing as do items presented earlier in the list. According to Atkinson and Shiffrin's (1968) dual storage model of memory, these terminal items are retrieved for the most part from a temporary short-term

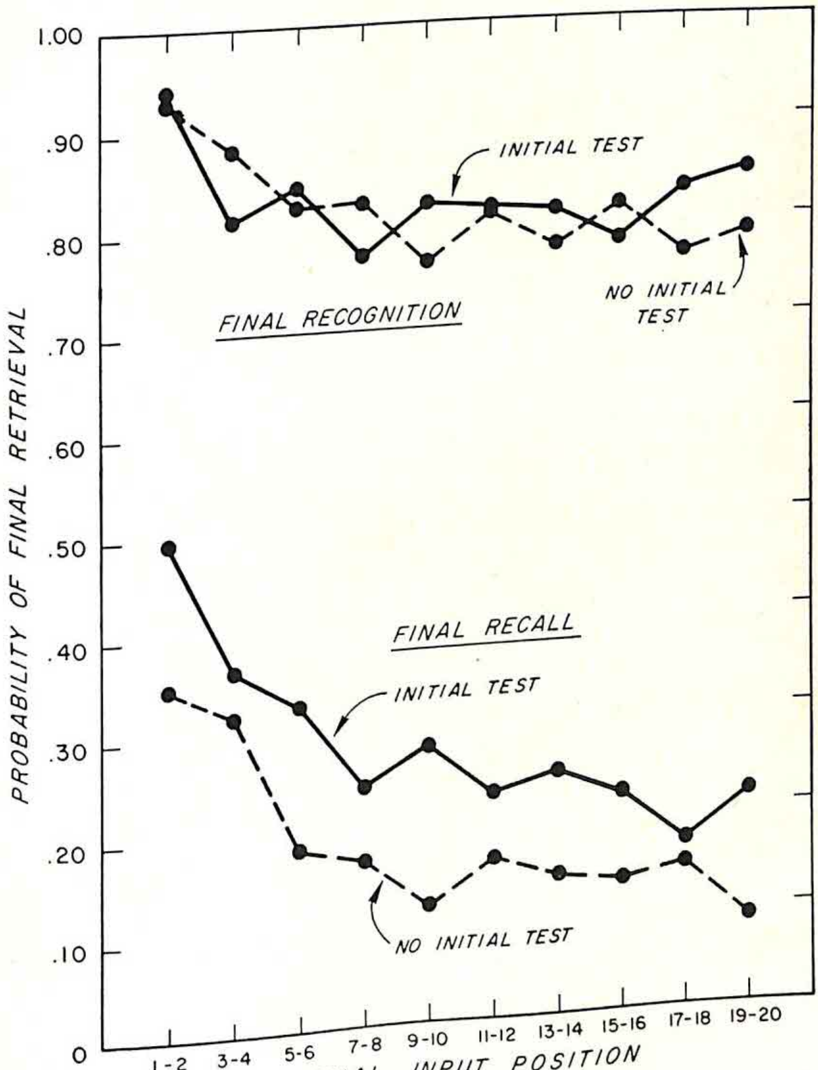

FIG. 2. The probability of final recall and final recognition as functions of an item's list position during study—for items from lists which receive an initial free recall test and those which receive no initial test.

store, while items presented earlier in a list are retrieved from a more permanent long-term store. It could be that the terminal items, being readily available for recall at the time of test, do not receive as much facilitation in final recall as do the items from earlier input positions. If such were the case, the final recall curves for items from initially tested and nontested lists would converge for the terminal positions. The data seem to argue against this possibility. Although it appears that the curves are converging for Positions 17–18, they diverge for Positions 19–20. The same curves were also plotted using only the final three tested lists and final three nontested lists. Although these curves are not shown here, they show no signs of converging over Positions 17–20. It appears that final retrieval is improved for all items from lists which receive an earlier recall test.

The results for final recognition are very different. There is no significant difference between the curves for tested (82.5%) and nontested (81.5%) lists, $t\ (23) = .56$, $p > .50$. It appears that items do not receive the same facilitation in recognition performance due to prior recall testing as occurs when they are tested for final recall. Tests were also performed to determine if the tested and nontested recognition curves showed either negative or positive recency. Linear trend analyses show that for these curves there is no significant decrease or increase over the terminal serial positions. For tested items, $F\ (1, 138) = 3.27$, and for nontested items, $F\ (1, 138) = .01$.

The probability that an item is retrieved from memory can be considered to be a function of both the availability of the item's trace in memory storage and the accessibility of that trace (Tulving & Pearlstone, 1966). A recognition test such as the one employed in the present study provides an excellent measure of the availability of items. Because *S*s are provided with the ultimate retrieval cues, the items themselves, the responses may involve only a comparison between the presented items on the forced-choice recognition test and the stored memory traces of items presented on the word lists. If prior testing of items produced any increment in the availability of these items in the memory store, this increment should be reflected in the curves for recognition of tested and nontested items in Fig. 2. In fact, there is no difference between the tested and nontested curves, indicating that prior testing of items has no effect on item availability.

On the other hand, there is a significant difference between the final recall curves for items from tested and nontested lists. Free recall requires the use of retrieval cues to gain access to stored information. Thus, success in retrieving an item on a free recall test is a function of the item's accessibility as well as its availability. Since the results for final recognition testing indicate that availability is not affected by prior testing, the results for final recall testing suggest that prior testing aids later retrieval by increasing the accessibility of items stored in memory.

Items that receive an initial test can be partitioned into items recalled on the immediate tests and those that were not recalled. Figure 3 presents the serial position curves for these types of items for both final recall and final recognition tests. Conditionalizing on prior recall undoubtedly introduces item selection effects, so interpretations of these curves must be made with care. It is interesting, however, to notice the low probability of final recall for items which were unrecalled initially. On the other hand, recognition performance on unrecalled items is considerably above chance. For recalled items that receive a final recall test, there is a definite negative recency effect. The trends in the other curves are less decisive and difficult to interpret due to either high or low performance levels and the previously mentioned selection effects.

When *S*s receive immediate recall tests they sometimes recall items from lists that were presented prior to the list being tested. Each of these prior list intrusions may have been initially presented in a list which received an immediate recall test or a list which did not. With regard to the original

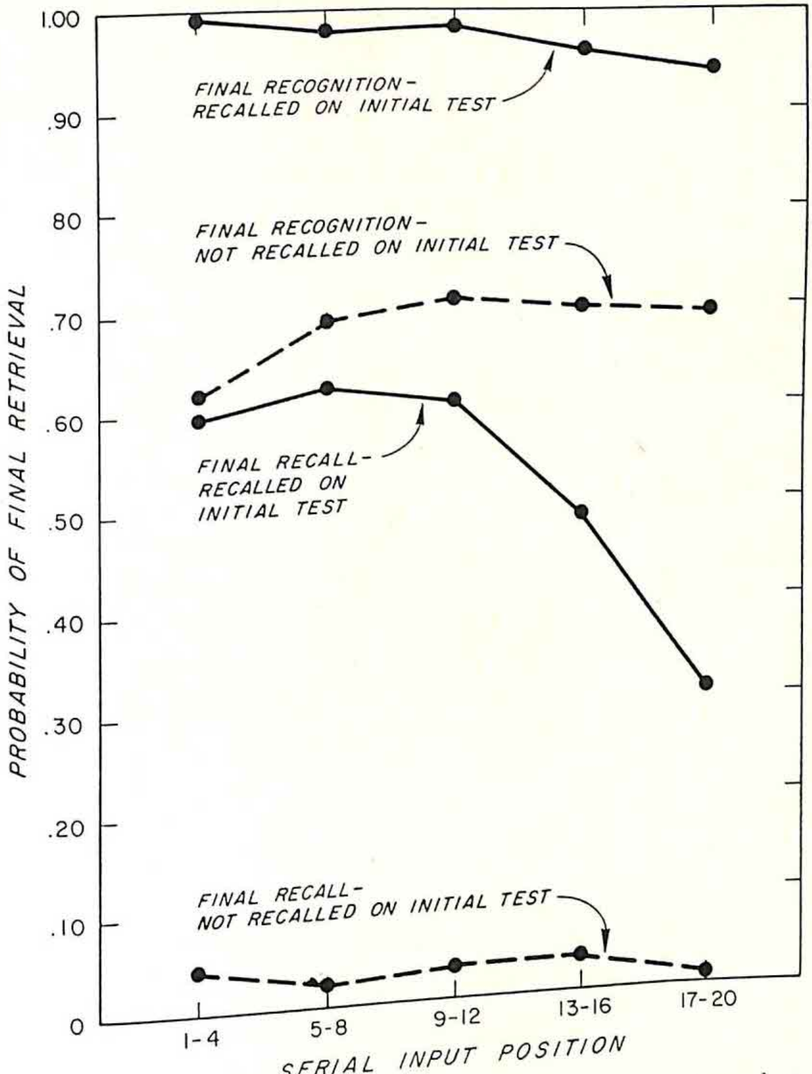

FIG. 3. The probability of final recall and recognition as functions of an item's list position during study, conditionalized on recall or failure to recall the item on the initial test.

serial positions of the intrusion items, the frequency with which these items occur decreases from a total of 31 for Serial Positions 1–4 to a total of 12 for Positions 17–20. Of the total number of intrusions, 81% are items from nontested lists, which indicates that if a list receives an initial test the items in that list are not likely to appear as intrusions on tests for other lists. So items from tested lists do not often become intrusion errors even though *S*s remember more of these items on a final recall test than they do items from nontested lists (see Fig. 2). It may be that *S*s label items in memory as having been tested or not and that this labeling affects the probability of an item becoming an intrusion error.

Craik (1970) found that the probability of an item being recalled on a final recall test was highly dependent upon the position at which it was recalled on the initial test. Items from early output positions in initial recall had the lowest probability of final recall, and those from final output positions had the highest probability of

final recall. Rundus et al. (1970) showed that the same is true for a 3-wk. delayed recognition test. However, as Rundus et al. point out, caution should be taken when interpreting this result because there is a substantial variability in the number of items different *S*s recall on the initial tests. As a result, *S*s do not contribute equally to all output positions. Rundus et al. corrected for this by Vincentizing their data, a procedure which divides each *S*'s output into equal parts so that the probability of retrieval on the final test can be plotted as a function of, for example, Output fifth. They found that the probability of 3-wk. recognition was still low for items output early in the initial recall, but instead of monotonically increasing the curve rose to a peak and dropped off for the final two-fifths of out-

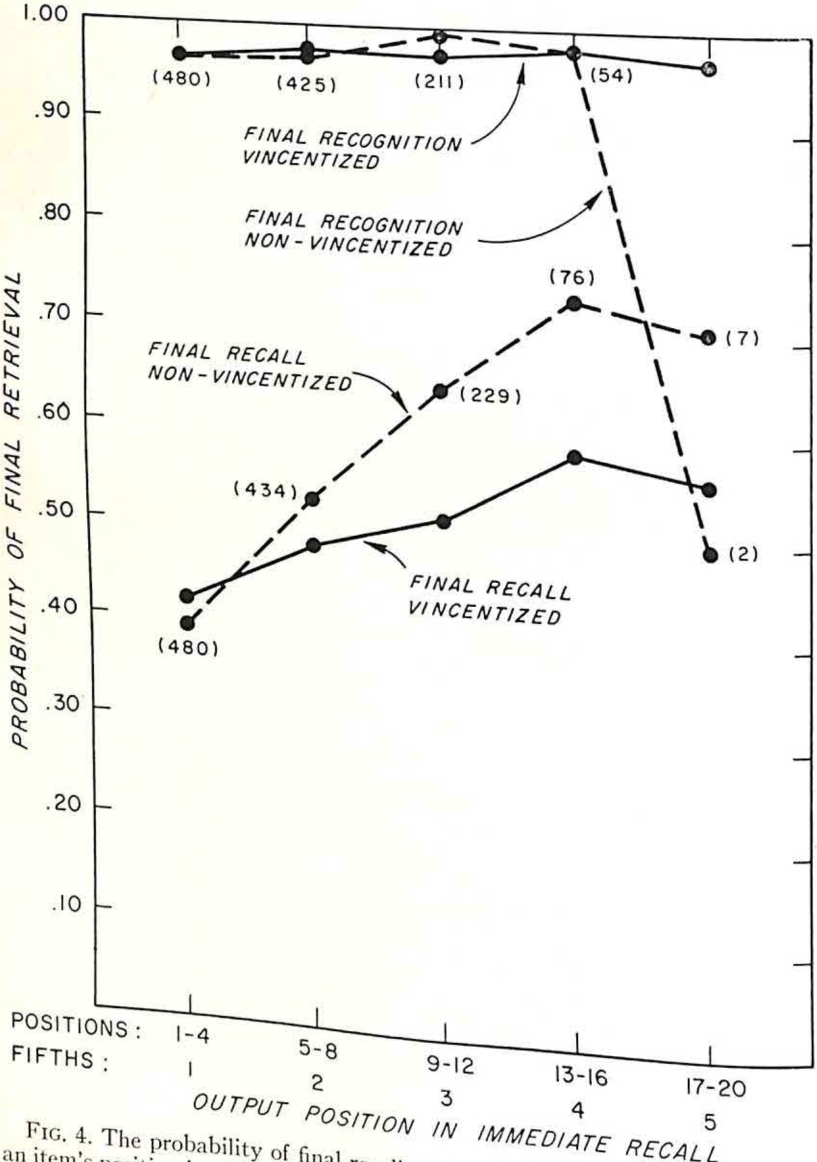

FIG. 4. The probability of final recall and recognition as functions of either an item's position in the initial recall output protocol (non-Vincentized) or the item's position in the Vincentized initial recall output protocol. (The number of items output in Positions 1–4, 5–8, 9–12, 13–16, and 17–20 in initial recall is affixed for non-Vincentized curves.)

put. In the present study non-Vincentized and Vincentized output curves were obtained for probability of final recall and final recognition. These curves are presented in Fig. 4.

All four curves show that performance is lowest for items from early output positions, increases for later positions, and decreases for final positions. It is doubtful that this decrease is significant for the non-Vincentized curves because the final points include data from only two items in the case of final recognition and seven items for final recall. A further problem in the interpretation of the data for final recognition performance for both Vincentized and non-Vincentized curves arises due to the high level of performance. The possibility of significant trends occurring over serial positions is virtually eliminated by this high performance level. The most interesting curve is that for Vincentized final recall. The decrease for final positions is not as pronounced as it was for recognition performance in the Rundus et al. (1970) study, and it only occurs for the final fifth of output. In fact, when immediate output protocols are divided into quartiles rather than fifths, this final recall curve shows no decrease for the final quartile.

The principal finding of this study is the higher probability of final recall of tested than of nontested items whereas no effect of prior testing was evident when the final test was a three-alternative forced-choice recognition test. This result is interpreted as evidence that prior free recall testing provides *S*s with easier access to the tested items but has no effect on item availability.

The existence of a statistically significant negative recency effect has been confirmed for final recall when the data from tested and nontested lists are combined. Although there was no significant decrease over terminal positions when tested and nontested items were considered separately for final recall and negative recency was not found when *S*s received a final recognition test, the theoretically important point is that in no instance was there positive recency on a final test. Cohen (1970) contends that his data demonstrate a positive recency effect for recognition testing, but the effect he obtained was small and no statistical test was reported. In light of the findings by Rundus et al. (1970) and McCabe and Madigan (1971) that a nonpositive recency effect is obtained when final testing is with a recognition test, the results from the present study cast some doubt upon the existence of a positive recency effect for final recognition testing.

REFERENCES

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*. Vol. 2. New York: Academic Press, 1968.

Cohen, R. L. Recency effects in long-term recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 672–678.

Craik, F. I. M. The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 143–148.

Craik, F. I. M., Gardiner, J. M., & Watkins, M. J. Further evidence for a negative recency effect in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 554–560.

Madigan, S. A., & McCabe, L. Perfect recall and total forgetting: A problem for models of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 101–106.

McCabe, L., & Madigan, S. A. Negative effects of recency in recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 307–310.

Rundus, D., & Atkinson, R. C. Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 99–105.

Rundus, D., Loftus, G. R., & Atkinson, R. C. Immediate free recall and three-week delayed recognition. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 684–688.

Tulving, E., & Pearlstone, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 1966, 5, 381–391.

(Received March 22, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 74–77

INFORMATION TRANSMISSION RATES IN A TASK REQUIRING MEMORY

HERBERT M. KAUFMAN
University of Connecticut

THOMAS J. HAMMELL
General Dynamics/Electric Boat Division, Groton, Connecticut

AND JERRY C. LAMB[1]
Naval Underwater Systems Center, New London, Connecticut

Information transmission rates for a serial processing task were investigated. The Ss were presented with a sequence of X's and O's on a computer-driven display and were required to respond to symbols either on the display (visible) or symbols removed from the display (remembered). Both the number of visible stimuli and the memory requirements were varied and S's information rate was measured. The results indicate that Ss used both parallel processing of visible stimuli and perceptual chunking of remembered stimuli to increase their transmission rate.

The most important characteristic of an information transmission channel is the rate (bits/time) at which it can transmit information. For the human processing visual information, a typical task is reading or typing text, i.e., reproducing a sequentially presented message. In the present study, the processing rate of humans was measured by using a simplified version of this task; the message was composed with an alphabet of two highly identifiable stimuli paired by a one-to-one code with key-pressing responses.

There are two categories of parameters involved in the visual processing task, one concerned with message characteristics (number of alternatives, redundancy, etc.) and the other with channel characteristics. In particular, the requirements for storage (memory) and amount of input message available at any given time (integration unit) are important. There have been many studies concerned with the effects of message parameters (e.g., Garner, 1962) on transmission rate but few on channel parameters. In the present experiment, message parameters were fixed while memory requirements and integration unit were varied.

[1] Requests for reprints should be sent to Jerry C. Lamb, c/o W. L. Clearwaters, Associate Director, Sensors, Naval Underwater Systems Center, New London Laboratory, New London, Connecticut 06320.

Shannon and Weaver (1949) proved that with proper encoding, the information capacity (in bits/time) of a channel can be reached even in the presence of noise. One method of achieving the capacity is to increase the length of the units of the message to be coded. In the present case, this corresponds to increasing the number of visible stimuli. When humans process visual information, as in reading, there is a considerable amount of integration of the text when many stimuli are available simultaneously for processing. This integration can be demonstrated easily by reading text through a small hole. The smaller the hole, or window, the slower the reading rate. Thus, in the present task, increasing the length of the message units should increase transmission rate.

If text is presented in acoustic form, as in speech or Morse code, the window size is necessarily very small. There can be no integration of the type discussed above, but there certainly is a considerable "backward" integration using stimuli stored in memory. This backward integration can be demonstrated easily by observing the lag between the time the symbol is heard (stimulus) and the time the message is recorded (response). This lag shows that much of the processing occurs with stored information; humans process the message from a short-term memory store rather than processing the message as it is received. In other words when no processing

constraints are imposed, *S*s tend to adopt strategies involving different amounts of memory load for each task to maximize transmission rate. The approach in the present study was somewhat different. With fixed message parameters, *S* was required to transmit the message under different memory loads, and thus a transmission rate could be measured for each condition.

The message was a string of X's and O's with only a portion of the total message visible at any given point in time. As *S* responded, these stimuli shifted across his field of view, as if the message was passing behind a window. Furthermore, the symbols that passed through and out of the window could be thought of as shifting into nonvisible locations consecutively further from the window. The memory requirements were controlled by having *S*s responding to both visible and nonvisible locations. For *S* to respond to a symbol occupying a nonvisible location, he had to remember the symbols that had shifted out of the window; the number of symbols he had to remember depended on the specified response location.

METHOD

Apparatus.—The apparatus consisted of a general purpose digital computer (DDP–24) with an associated cathode ray tube (CRT) display and typewriter keyboard. The CRT presented the stimuli to *S*, who responded by pressing an appropriate key on the keyboard. When the response was made, the computer immediately updated the display, so that the rate of information presentation was under *S*'s control. Following each response, the computer automatically recorded the correct stimulus as well as *S*'s response and reaction time (time since the previous response).

Stimuli.—The stimuli displayed on the CRT consisted of a series of binary symbols (× and ○) presented in random order, each with a probability of occurrence equal to one-half.

Subjects.—The *S*s were sixteen men and two women employees of the Research and Development Department of General Dynamics, Electric Boat Division. Their ages ranged from 23 to 45 yr. None of the *S*s had practice with the task prior to the experiment.

Procedure.—The number of symbols displayed simultaneously on the CRT was called window size; window sizes of one, two, and four were used (designated as W_1, W_2, and W_4).

The *S*'s task was to respond to the symbol occupying a specified location: (*a*) the left-most location in the window ("visible" response location, designated as VIS); (*b*) the first nonvisible location to the left of the window ("memory-1" response location, M_1); (*c*) the fourth nonvisible location to the left of the window ("memory-4" response location, M_4).

In the latter two response locations, the stimuli were not visible at the moment of response. Therefore, the stimuli had to be remembered from the time they were visible in the window.

The experimental design consisted of the factorial combination of three window sizes and three response locations, which yielded nine experimental conditions. The information load (two equiprobable stimuli yielding 1 bit per symbol) of the message was constant across conditions.

Two *S*s were randomly assigned to each of the nine conditions; each *S* participated in 12 sessions, 2 per day. A session consisted of a continuous sequence of 120 trials. To minimize the effects of starting and stopping, data from the middle 100 trials only of each session were analyzed. The task was to press a key corresponding to the symbol in the response location corresponding to that particular experimental condition; *S*s were told to respond as quickly and accurately as possible.

Within a few milliseconds after a key-pressing response, the display was updated. All but the left-most visible symbol was shifted one location to the left, and the left-most symbol was removed from the CRT. This procedure can be thought of as shifting the left-most symbol into the first nonvisible location to the left of the window. At the same time, a randomly selected symbol was entered at the right-most location in the window.

RESULTS

The dependent measures were the amount of time taken to complete the 100 trials and the number of errors made. From these, average response rate (symbols/second) and average information transmission rate (bits/second) were calculated. Figure 1 shows the relation between the independent variables and information transmission rate for the last experimental session. The results are similar for the number of symbols/second data except that the point for VIS–W_1 is relatively lower for information transmission because of errors made by *S*.

An analysis of variance conducted on the information transmission measure showed significant effects of both window size, F (2, 9) = 11.7, $p < .005$, and response location, F (2, 9) = 11.40, $p < .005$, but no significant interaction. A Scheffé range test confirmed that all levels of both

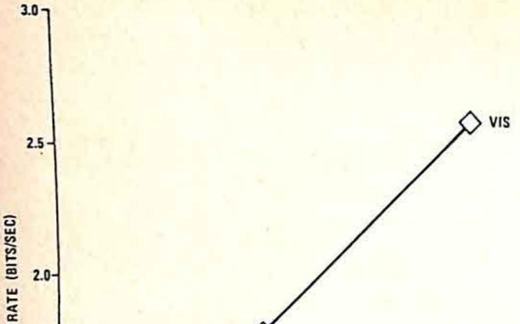

FIG. 1. Window size versus average information transmission rate for the final session.

variables, except VIS versus M_1 and W_1 versus W_2, were significantly different ($p < .05$) from each other.

As memory load increased, information transmission rate decreased. The effect was consistent and approximately equal for all window sizes, except that there was no difference between the VIS–W_1 and the M_1–W_1 conditions.

As window size increased, the transmission rate increased at the same rate regardless of the memory load imposed on *S* (no significant interaction between memory load and window size). This constant rate increase indicates that *S*s were able to use all the available stimuli effectively and independently of memory load. In fact, the most striking finding of this study was that *S*s in the M_4 condition, when given a sufficient number of visible stimuli, increased their transmission rate to approximately the same level as that obtained with a much lower memory load and either one or two visible stimuli (VIS–W_1, M_1–W_1, and M_1–W_2).

Discussion

The effect of increasing window size was to increase information transmission rate, which indicates that *S* was processing several stimuli at the same time (parallel processing). The most dramatic demonstration of the effectiveness of this mechanism was in the VIS-W_1 condition, where the transmission rate was much lower than a straightforward extrapolation of the transmission functions from the other data points would imply. The VIS-W_1 condition was the only one in which no parallel processing could occur; all other conditions had at least one other stimulus that either was visible or available in memory at the time the response was made.

Two stimuli needed to be available for parallel processing to take place. In the case where the second stimulus was available only in memory (window size of one), the advantage of parallel processing offset the memory decrement. For window sizes greater than one, adding visible stimuli increased the transmission rate, whereas adding to the memory requirements decreased the transmission rate.

As noted, the transmission rate in one of the M-4 conditions (M_4–W_4) was as high as a condition with no memory requirements (VIS-W_1). The question arises as to how this relatively high level of performance was achieved. When questioned, *S*s reported that they used the following perceptual coding technique: with the W_4–M_4 condition, the visible symbols were tagged as a "chunk" of four, an implicit perceptual configuration such as ××○×, ○××○, etc. This chunk was stored in memory while the next visible group of four was being encoded and stored. After the second chunk was stored, the first group was reported. The next visible group of four was then chunked, and so on.

The chunking sequence was as follows (Fig. 2): (*a*) Group 1 (of four visible symbols) was coded and stored in memory as a chunk. Four responses from a previous group were made, thus advancing from Trial 1 to Trial 5. Group 2 (of four visible symbols) was coded and stored in memory as a chunk. At this time two chunks (i.e., Groups 1 and 2) were stored in memory. Group 1 was now decoded and reported by four rapid responses. (*b*)

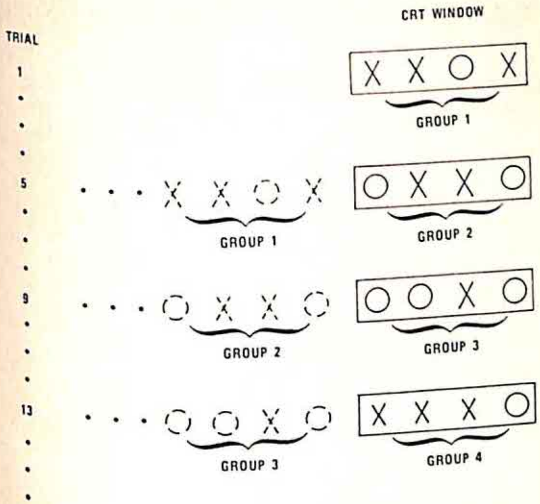

FIG. 2. Illustrative example of the chunking sequence reported by *S*s in the M4–W4 condition.

Group 3 (or four visible symbols) was coded and stored in memory as a chunk. Again, two chunks (Groups 2 and 3) were stored in memory. (*c*) Group 2 was now decoded and reported by four rapid responses, thus advancing from Trial 9 to Trial 13. (*d*) Group 4 (of four visible symbols) was coded and stored in memory as a chunk, and so on.

By chunking, *S*s may have been able to reduce the number of discrete quantities they had to store in memory for this condition from five to two.

A similar coding scheme was used by *S*s in the M_4–W_2 condition. However, because of the smaller window size, the scheme was less efficient than that for M_4–W_4. With a window size of two, the chunk of a visible group contained two symbols; therefore, even under ideal conditions, the memory load was three chunks. This factor accounted for the drop in efficiency. With a window size of one, there was no opportunity to chunk visible stimuli, and therefore all symbols had to be stored.

The coding procedure, used without instruction by *S*s, was very similar to the chunking code reported by Miller (1956). In the present study, however, the coding was perceptual (implicit) rather than learned (explicit).

There was behavioral evidence to support *S*'s reports of coding. In the M_4–W_4 condition, *S*s showed a response time pattern of groups of four quick responses separated by long intervals. Earlier studies (Glanzer & Fleishman, 1967; Kleinberg & Kaufman, 1971; Lamb & Kaufman, 1966) found that coding takes time. The response pattern of the *S*s suggests that they were coding and storing information during the long pauses and then responding quickly with previously stored information. Further support is provided by the fact that *S*s in the M_1 and VIS conditions where coding was not feasible showed no such response patterns. The nature of this perceptual coding process is presently the subject of further experimentation.

In the high memory load condition, the effect of increasing window size is to make available to *S* stimuli that can be encoded into chunks, effectively reducing memory load. Although this encoding takes time, the reduction in storage and retrieval load for individual stimuli results in an overall processing gain.

The results show two characteristics of the processing of sequentially presented visual information. First, all visible stimuli, up to the four used in this study, are processed concurrently so that the transmission rate is monotonic with the amount of message visible. Second, when the message must be routed through a memory store, *S*s resort to an implicit coding (perceptual chunking) to reduce the number of units in storage.

REFERENCES

GARNER, W. R. *Uncertainty and structure as psychological concepts*. New York: Wiley, 1962.

GLANZER, M., & FLEISHMAN, J. The effect of encoding training on perceptual recall. *Perception & Psychophysics*, 1967, **2**, 561–564.

LAMB, J. C., & KAUFMAN, H. M. The effects of coding procedure on human information transmission, *Perception & Psychophysics*, 1967, **2**, 256–262.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information, *Psychological Review*, 1956, **63**, 81–97.

KLEINBERG, J., & KAUFMAN, H. Constancy in short-term memory: Bits and chunks, *Journal of Experimental Psychology*, 1971, **90**, 326–333.

SHANNON, C. R., & WEAVER, C. R. *The mathematical theory of communication*, Urbana: University of Illinois Press, 1949).

(Received March 29, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 78–84

ROLE OF COLLATERAL BEHAVIOR IN TEMPORAL DISCRIMINATION PERFORMANCE AND LEARNING IN RATS [1]

HOWARD GLAZER AND DEVENDRA SINGH [2]

University of Texas at Austin

Two studies were designed to test the functional significance of collateral behavior during the acquisition of a temporal discrimination (differential reinforcement of low response rates, DRL, 10-sec. schedule). In the first study, three groups of rats were trained, each under a different condition of body restriction, and a fourth group served as a stress control. All except completely restricted *S*s showed DRL acquisition. In the second study, *S*s were placed on a DRL 10-sec. schedule, one group under conditions of complete restriction, the other group under no restriction conditions, following which half of each group was switched to the other restriction condition. The *S*s switched from restriction to no restriction showed rapid DRL acquisition, while *S*s switched in the opposite direction showed a decrement in DRL performance. On the basis of these results, it was concluded that rats are capable of learning temporal contingencies while deprived of gross motoric behavior but are not able to perform appropriately under such conditions.

Skinner (1948) was the first to report that in a situation where reinforcement is delivered at fixed intervals and is not contingent on any specific behavior, animals develop highly stereotyped patterns of responding which he termed "superstitious behavior." Such stereotyped behavior has also been noted in temporal discrimination tasks where reinforcement is contingent on a specific terminal response. For example, animals trained on a schedule of differential reinforcement of low response rates (DRL) develop highly stereotyped idiosyncratic behavior which fills the intervals between occurrences of the designated operant (cf. Holtz & Azrin, 1963; Wilson & Keller, 1953). The role of such stereotyped or collateral behavior in DRL performance is evident in studies by Laties, Weiss, Clark, and Reynolds (1965), and Laties, Weiss, and Weiss (1969). These authors found that well-developed DRL performance could be disrupted by blocking the collateral behavior; when the collateral behavior was again permitted, the reinforcement returned to its original level.

The most frequent functional role ascribed to these collateral behaviors is as "timing behavior," the idea being that they develop to bridge the time gaps separating reinforcements. It has been suggested that the occurrence of collateral behavior increases the probability of reinforcement and thus tends to persist. The mere occurrence of collateral behavior does not, however, mean that such behavior is necessary or sufficient for the acquisition of a temporal discrimination. As Anger (1963) has suggested, one way to determine the functional significance of these collateral behaviors is to systematically manipulate them and observe their effects on the acquisition and performance of a temporal discrimination. If the development of collateral behavior is critical for temporal discrimination learning, animals unable to engage in overt collateral behavior should not be able to learn the temporal discrimination.

To test these relationships, two experiments were conducted. In the first, rats were trained on a DRL schedule under conditions of differential bodily restriction. The second experiment was conducted to ascertain whether observed DRL per-

[1] This research was supported by University Research Institute Project R-130 14-7499-0130 awarded to Devendra Singh and by National Science Foundation Grant GB-14990 X awarded to Abram Amsel, whom the authors wish to thank for a critical reading of the manuscript.

[2] Requests for reprints should be sent to Devendra Singh, Department of Psychology, University of Texas at Austin, Austin, Texas 78712.

formance deficits produced by restriction of bodily movement in Exp. I reflected a learning or performance deficit.

Experiment I

Method

Subjects.—Twenty-four male Holtzman albino rats approximately 100 days old served as *Ss*. Each *S* was housed in an individual cage and was provided with ad-lib water throughout the experiment. A 12-hr. dark-and-light cycle was maintained.

Apparatus.—Two identical Plexiglas operant conditioning chambers (21.25 × 16.25 × 37.50 cm.) and three restraining boxes were used. In each chamber, a 2 cm. long × 1 cm. wide lever was mounted on the center of the end panel, 11.25 cm. above floor level. An upward pressure of 10 gm. on the lever activated the food dispenser. Directly below the lever, a food cup was mounted on the floor. Each chamber was housed in a sound-attenuated box equipped with a ventilation fan.

The three restraining boxes designed for complete restriction, partial restriction, and stress control were made of .63-cm. plywood. The complete and partial restriction boxes measured 20.00 cm. long × 13.25 cm. high × 8.75 cm. wide. On one end of these boxes was a hinged door with a 3.00-cm. circular hole in its center. The complete restriction box was also fitted with lateral partitions on either side of the head hole and a plywood insert which could be clamped down by bolts from the top of the box and which covered the length and width of the box. The stress control box measured 5.00 cm. long × 13.25 cm. high × 8.75 cm. wide. It was completely opened at both ends and was fitted with a plywood insert like that described in the complete restriction box. In the complete restriction box, *S* was clamped tightly with his head protruding through the head hole which restricted *S* to vertical head movement only. The *Ss* in the partial restriction box could move their bodies inside the box to a very limited degree, and the head protruding through the head hole could move in any direction. The stress control box was placed around the middle of *S*, allowing complete freedom of limb and head movement. Behavioral recording and event control was done with solid-state circuitry.

Procedure.—The *Ss* were randomly divided into four groups of six each: complete restriction, partial restriction, nonrestriction, and stress control. After 5 days of handling, all *Ss* were place on a 10 gm/23 hr deprivation schedule and maintained on this schedule throughout the experiment. One week following the introduction of the deprivation schedule, *Ss* were adapted to their respective restriction boxes for ½ hr. a day for 3 days. Box adaptation consisted of securing *S* in the appropriate restriction box and placing the box on an empty table top. The *Ss* in the nonrestriction group were allowed to investigate the table top for ½ hr. Following restriction-box adaptation, the operant conditioning phase was started. The *Ss* from the complete restriction, partial restriction, and stress control groups were secured in their respective restriction boxes and were positioned in the operant chamber in such a manner that raising the head resulted in a lever press and lowering the head permitted access to the food cup. The *Ss* from the nonrestriction group were placed in the operant chamber unrestrained. All *Ss* were given 25 reinforcements (45 mg.-Noyes pellet) a day on a Fixed Ratio (FR 1) schedule unde rtheir respective restriction conditions for 2 days. Then the *Ss* were shifted to a DRL 10-sec. schedule and trained for ½ hr. a day for 20 consecutive days. Finally, all *Ss* were tested for extinction for ½ hr. a day for 5 consectuive days. Throughout the experiment, four within-group random running orders were employed. Running orders were randomized over days but held constant between groups within days. Number of responses over the ½-hr. sessions were recorded by a series of ring counters in 5 successive 2-sec. time intervals (timed from the previous response) for each *S*.

Results

Interresponse time (IRT) data for each *S* for all days of acquisition were converted into interresponse time per opportunity to respond (IRT/OP.) scores (Anger, 1963). Percent reinforcement for each *S* was calculated using the formula:

$$\frac{\text{number of reinforced responses}}{\text{total number of responses}} \times 100.$$

Percent of theoretically available reinforcements obtained was calculated using the formula:

$$\frac{\text{number of reinforcements}}{180} \times 100.$$

Figure 1 presents the percent reinforcement over 20 days of acquisition for each group and IRT/OP. analysis for acquisition on Days 1, 10, and 20 for each group. As the figure indicates, performance of the complete restriction group is lower than that of the other groups and, unlike the others, shows no progressive improvement as training continues. Analysis of variance of the percent reinforcement data yielded significant effects of group, F (3, 20) = 29.08; days, F (19, 381) = 15.20; and Group × Days, F (57, 381) = 11.70, all $ps < .01$, indicating a significant increase in percent reinforcement over days for all except the complete restriction group.

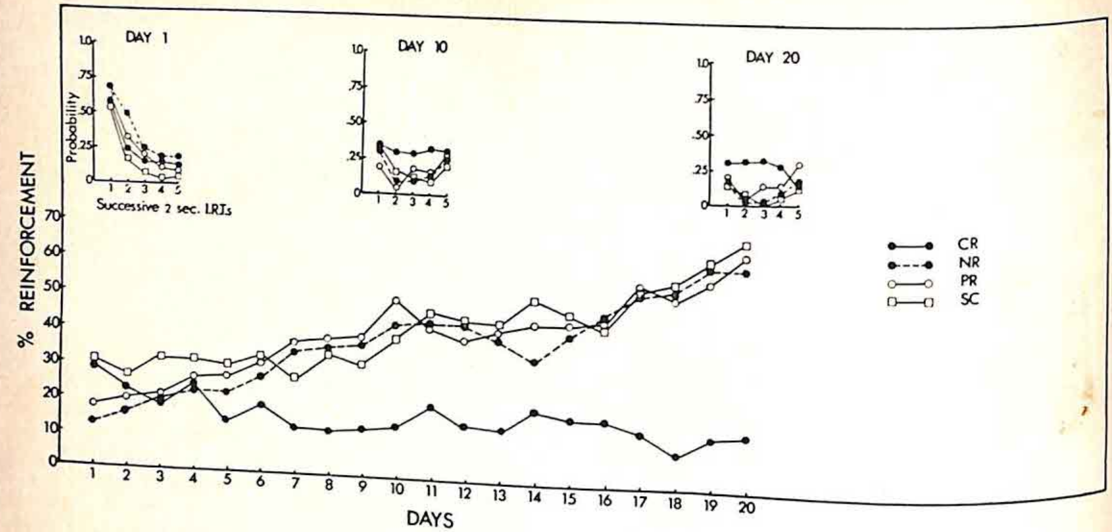

FIG. 1. Percent reinforcement for each group over 20 half-hour daily sessions, IRT/OP. curves for Days 1, 10, and 20.

The inability of the complete restriction group to learn the temporal discrimination becomes clear from the inspection of the IRT/OP. curves (Fig. 1). While the other groups show a bimodal distribution which is present by Day 10 and becomes more pronounced by Day 20, further suggesting a progressive efficiency of discrimination, no such trend is evident for the complete restricted group.

In order to determine whether the increase in percent reinforcement, observed in all but the complete restriction group, was indeed indicative of an increased temporal discrimination efficiency, the percent of theoretically available reinforcements obtained was also calculated. Table 1 shows the percent of theoretically available reinforcements obtained for each group over the 20 days of acquisition. As the table shows, all except the completely restricted group obtained better than 60% of the available reinforcements by the termination of acquisition. An analysis of variance indicated significant effects of group, days, and Group × Days, F (3, 20) = 17.26; F (19, 381) = 12.37; and F (57, 381) = 4.73, all ps < .01, respectively.

An examination was also made of total frequency of responding over daily sessions to determine whether the performance deficit of the complete restriction group resulted from a higher response rate due to stress or general arousal. Analysis of variance did not reveal a significant group main effect, F (3, 20) = 3.00, $p > .05$, but did yield a significant Group × Days interaction, F (57, 381) = 8.00, $p < .01$. This

TABLE 1
PERCENT OF THEORETICALLY AVAILABLE REINFORCEMENTS OBTAINED (EXPERIMENT I)

Group	Day 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
NR	19	31	35	39	39	38	40	43	47	44	48	50	49	50	52	56	60	59	59	63
PR	18	33	37	41	40	40	46	45	50	52	49	48	47	49	52	54	59	60	57	62
CR	14	14	15	14	17	16	18	15	22	16	17	18	15	21	20	19	21	22	17	20
SC	10	15	19	21	29	31	30	32	38	36	35	40	39	48	41	49	52	55	58	61

Note.—Groups are abbreviated as follows: no restriction = NR; partial restriction = PR; complete restriction = CR; stress control = SC.

significant interaction resulted from an increase in total responses for the complete restriction group and a decrease in the total responses for the remaining groups over the first 3 days. The groups showed no differences for Days 4–20 with respect to total number of responses.

Finally, an analysis of the data from total response extinction showed a significant days effect, F (4, 81) = 9.84, $p < .01$, indicating that all groups make fewer and fewer responses as extinction continues but that the groups do not show differential rates of extinction.

Discussion

Deprived of overt collateral behavior, the complete restriction group showed no evidence of having learned a temporal discrimination over 20 half-hour sessions, as reflected in the acquisition of DRL performance. The fact that the partial restriction and stress control groups did not differ significantly from the nonrestriction group suggests that complete freedom of bodily movement is not essential for the acquisition of DRL performance; a small amount of overt behavior appears sufficient. On the basis of the percent reinforced responses and IRT/OP. data, it is possible that the groups which did appear to acquire the DRL schedule had not learned the temporal discrimination but were simply responding very slowly, for example, with 39- or 40-sec. latencies. The percent of available reinforcement obtained data, however, makes it clear that this is not the case. In order to obtain 60% of the available reinforcements, *S*s' responses could not have averaged greater than about 15 sec. in latency, indicating that they had indeed learned the temporal discimination to some extent.

It should be noted that the complete restriction *S*s were effectively deprived of all overt behavior, since the only overt behavior which was permitted (vertical head movement) resulted in a lever response and was thus directly incompatible with learning the required DRL behavior. However, the analysis of total frequency of responding suggests that this was not the factor that retarded DRL acquisition, at least not after the first few days. There is, of course, the possibility that failure of complete restriction *S*s to learn can be attributed to stress caused by body restraint. However, this is not very likely in light of the finding of no difference between the stress control and nonrestricted *S*s, and on the basis of the findings of Perhach and Barry (1970). Using increase in plasma corticosterone as an index of stress, Perhach and Barry found that bodily restraint in rats does not produce an increase in plasma corticosterone level within 24 hr. after restraint.

Finally, it appears that the differential bodily restriction had no affect on the rate of the response decrement during extinction.

Experiment II

The findings of Exp. I clearly show that absence of overt collateral behavior and failure to acquire a DRL response are correlated. Still, the possibility exists that the inability of the complete restriction group to acquire a DRL response reflects the operation of a performance rather than a learning deficit. It can be argued that a generalized nonassociative arousal factor potentiates the only available behavior (vertical head movement), which is incompatible with DRL acquisition. Simply because the performance of the completely restricted group does not reflect learning of a temporal discrimination does not mean it has not been learned. That is to say, it is possible that the completely restricted group did learn some type of covert timing behavior but that some performance variable such as generalized arousal rendered them incapable of performing appropriately under bodily restriction conditions. In order to determine the validity of this possibility, a second experiment was conducted.

Method

Subjects.—The *S*s were 16 male Holtzman rats approximately 100 days of age. All *S*s were maintained under conditions identical to those described in Exp. I.

Apparatus and procedure.—The apparatus, handling, and deprivation schedules were the same as in Exp. I.

The study was a 2 × 2 factorial involving two phases. In Phase 1, half of the *S*s were completely restricted and the other half not restricted. In Phase 2, half of the *S*s in each of the two groups were switched to the opposite condition, while the other half remained under their original acquisition condition. This yielded four groups of four *S*s each: restricted–restricted, restricted–nonrestricted, nonrestricted–nonrestricted, and nonrestricted–restricted. For 3 days prior to acquisition and for 3

days intervening between Phases 1 and 2, *S*s which were to be under the restricted condition were adapted to complete restriction for ½ hr. per day as in Exp. I. Following restriction-box adaptation, all *S*s were placed in the operant conditioning chamber under their respective Phase 1 restriction conditions with no reinforcement available, so that a base-line operant rate could be established under the two restriction conditions. Operant levels were recorded for ½-hr. sessions each day for 2 days. This was followed by 2 days of 25 FR 1 trials for each *S* under their respective restriction conditions. Then all *S*s were trained for ½ hr. a day for 20 consecutive days on a DRL 10-sec. schedule. Following Phase 1 training was 10 days of Phase 2 training under appropriate conditions of restriction, followed in turn by 5 days of extinction. All the sessions for Phase 1 and Phase 2 and extinction were of ½-hr. duration and data were collected as in Exp. I.

Results and Discussion

Analysis of operant level performance showed a significant difference in mean number of responses between restricted and nonrestricted groups on the first day (4.7 vs. 1.2 responses), t (14) = 1.89, $p < .05$; however, no significant difference between groups was evident on the second day (1.4 vs. 1.1), indicating that the completely restricted *S*s habituate rapidly.

The mean percent reinforcement data for Phases 1 and 2 for each group are presented in Fig. 2, which also shows IRT/OP. analysis for Days 1, 10, 20, 25, and 30 for each group. Table 2 shows the percent of theoretically available reinforcements obtained for each group over the two phases. Phase 1 results simply confirm the findings of Exp. I. The percent reinforcement analysis of variance showed significant effects of Phase 1 restriction condition, day, and Day × Phase 1 Restriction Condition, F (1, 12) = 25.60; F (19, 228) = 7.10; and F (19, 228) = 2.70, all ps $< .01$, respectively. The percent of theoretically available reinforcement analyses also yielded significant effects of Phase 1 restriction condition, F (1, 12) = 14.21; day, F (19, 228) = 9.67; and Day × Phase 1 Restriction Condition, F (19, 228) = 4.97, all ps $< .01$. Phase 1 frequency of responding analysis showed no significant effects. The IRT/OP. analysis is further confirmation that as training progresses, the unrestricted groups develop

TABLE 2

Percent of Theoretically Available Reinforcements Obtained (Experiment II)

Group	Day 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
RR	10	14	17	20	21	19	22	18	21	23	19	21	17	14	17
RN	8	15	12	18	22	20	23	19	21	25	21	20	17	13	19
NN	12	21	28	30	39	35	42	39	41	43	46	45	54	63	59
NR	14	21	28	34	41	39	40	37	39	42	47	44	52	61	59

Group	Day 16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
RR	15	20	12	17	20	24	21	17	24	20	23	21	15	17	20
RN	15	22	14	17	21	55	52	60	58	54	56	52	59	57	61
NN	62	65	60	63	69	66	64	67	63	62	60	66	70	65	72
NR	62	65	60	63	60	38	34	25	21	18	22	19	24	20	18

Note.—Groups are defined as follows: restriction–restriction = RR; restriction–nonrestriction = RN; nonrestriction–nonrestriction = NN; nonrestriction–restriction = NR.

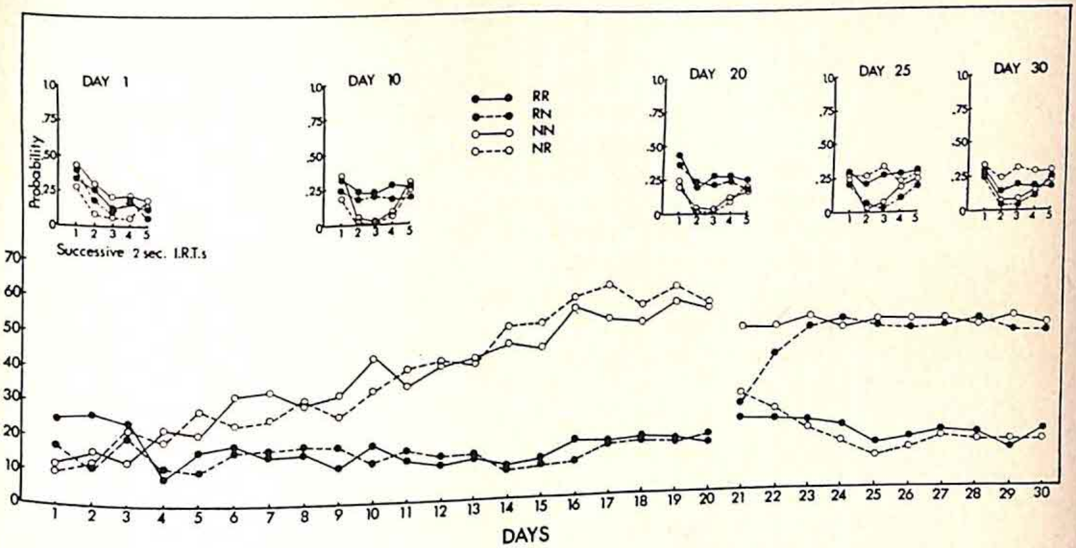

FIG. 2. Percent reinforcement for each group for Phase 1 (Days 1–20) and Phase 2 (Days 21–30), and IRT/OP. curves for Days 1, 10, 20, 25, and 30.

performance appropriate to the DLR schedule, while the restricted groups do not.

Phase 2 (Days 21–30) percent reinforcement performance shows significant effects of Phase 2 restriction, F (1, 12) = 22.43; Days × Phase 2 Restriction, F (9, 108) = 2.40; and Days × Phase 1 Restriction × Phase 2 Restriction, F (9, 108) = 3.00, all ps < .01. While it took 3 days for the percent reinforced responses of the restriction–nonrestriction group to reach the level of the nonrestriction–nonrestriction group in Phase 2, a much more abrupt change was observed in the percent of theoretically available reinforcements obtained for the restriction–nonrestriction groups. Phase 2 percent of theoretically available reinforcement obtained analysis showed significant effects of Phase 2 restriction only, F (1, 12) = 31.37, p < .01. In Phase 2, total response data showed no significant effects. The groups switched from nonrestriction to restriction and from restriction to nonrestriction showed drastically changed DRL performance while, of course, the restriction–restriction and nonrestriction–nonrestriction groups showed no change. The terminal performance of the nonrestriction–restriction group in Phase 2 was about the same as the level of the restricted groups in Phase 1 and of the restricted–restricted group in Phase 2. The decrement in temporal discrimination of Group nonrestriction–restriction is also reflected in the flattening of the IRT/OP. curve and the abrupt decrease in the percent of theoretically available reinforcement obtained. On the other hand, the DRL performance of the restriction–nonrestriction group changed to the level of Group nonrestriction–nonrestriction in Phase 2, and this level of performance was sustained for the remainder of training. This was the case for both percent reinforcement (by Day 23), for the IRT/OP. distribution (by Day 25), and for the percent theoretically available reinforcement obtained (immediately).

The deterioration of performance in the nonrestriction–restriction group is in accord with the findings of Laties et al. (1969) that interference with collateral behavior disrupts a learned temporal discrimination. However, the rapidity of acquisition of DRL performance shown by the restriction–nonrestriction group in Phase 2, as compared to the DRL acquisition rates of Groups nonrestriction–nonrestriction and nonrestriction–restriction in Phase 1, implies that preventing collateral behavior in Phase 1 did not entirely eliminate the possibility of learning a temporal

discrimination. It should be noted that the differences between nonrestriction–restriction and restriction–nonrestriction in Phase 2 were in percent reinforcement, percent theoretically available reinforcements obtained, and the IRT/OP. distribution, not in total number of responses per day either in acquisition or in extinction, although all groups made significantly fewer responses as extinction progresses, F (4, 49) = 7.32, $p < .01$. Again differential bodily restriction had no effect on the rate of response decrement during extinction.

General Discussion

Taken together, findings from these two experiments suggest that *S*s can learn a temporal discrimination under conditions of extreme restriction of bodily movement, but that they are unable to perform appropriately under these conditions. Furthermore, the observed disruption of DRL performance cannot, for several reasons, be adequately explained in terms of restriction-produced increases in arousal giving rise to behavior incompatible with DRL performance (i.e., high response rate), First, in neither experiment did restricted *S*s show a greater number of responses per session either in acquisition or extinction. Second, the operant level data from Exp. II show no significant differences between restricted and nonrestricted groups after the first day. Finally, a stress control group performed as well as the nonrestricted group.

The most intriguing question which arises from the present findings is why *S* shows a performance deficit but little if any learning deficit when restricted in bodily movement. Obviously, the complete restriction condition does not prevent *S* from engaging in behavior at molecular level or from developing other mediational behavior. The fact that *S*s can, to some extent, learn a temporal discrimination under the complete restriction conditions suggests that these *S*s were using some fractional response or mediational behavior as a cue to bridge the time interval. In this case, the question becomes why *S* is unable to use these cues —whatever their nature—during performance when transferred to restriction.

One may argue that under complete restriction, the only gross response available to the animal is directly incompatible with the designated operant, and therefore *S* makes more errors and the resultant poor performance. This argument, if anything, creates more problems. First, such an argument has to imply that an *S* under complete restriction cannot learn the temporal discrimination; the performance of the group switched from restriction to nonrestriction indicates otherwise. Second, it will also imply that *S* has to use a gross motor response to bridge a time interval and that some fractional or mediational behavior cannot serve as a functional cue for performance on a learned temporal discrimination task. Finally, if one assumes that *S*s, while learning the temporal discrimination, use fractional or mediational behavior but need some gross response for adequate performance, it is difficult to comprehend the utility and mechanics of such a substitution of cues.

REFERENCES

Anger, D. The role of temporal discrimination in the reinforcement of Sidman Avoidance Behavior. *Journal of Experimental Analysis of Behavior*, 1963, **6**, 477–506.

Holtz, W. C., & Azrin, N. H. A comparison of several procedures for eliminating behavior. *Journal of Experimental Analysis of Behavior*, 1963, **6**, 281–290.

Laties, V. C., Weiss, B., Clark, R. L., & Reynolds, M. D. Overt "mediating" behavior and the discrimination of time. *Journal of Experimental Analysis of Behavior*, 1965, **8**, 107–116.

Laties, V. C., Weiss, B., & Weiss, A. B. Further observations on overt "mediating" behavior and the discrimination of time. *Journal of Experimental Analysis of Behavior*, 1969, **12**, 43–57.

Perhach, J. L., & Barry, H. Stress response of rats to acute body or neck restraint. *Physiology and Behavior*, 1970, **4**, 443–447.

Skinner, B. F. "Superstition" in the pigeon. *Journal of Experimental Psychology*, 1948, **38**, 168–172.

Wilson, M. P., & Keller, F. S. On the selective reinforcement of spaced responding. *Journal of Comparative and Physiological Psychology*, 1953, **46**, 190–193.

(Received April 2, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 85–92

EAR DIFFERENCES AND DELAYED AUDITORY FEEDBACK: EFFECTS ON A SPEECH AND A MUSIC TASK[1]

JOHN L. BRADSHAW, NORMAN C. NETTLETON, AND GINA GEFFEN

Monash University

Left–right differences between ears were studied employing delayed auditory feedback (DAF). The *S*s were timed while reading prose passages and word lists and while piano playing. Reading times were found to be significantly lengthened with disruptive DAF to the right ear, and playing times were similarly increased when the DAF was played to the left ear, under certain conditions of dichotic stimulation. These effects occurred when a secondary input to the other ear consisted either of long-delayed nondisruptive feedback, or instantaneous amplified or attenuated feedback, but were not apparent with constant-level 90-db. white noise. The need for competitive stimulation in order to obtain left–right differences was considered, together with the possible contributions of handedness, cerebral asymmetry, ipsilateral pathways, and transcommissural routes. The findings were also discussed in terms of attentional processes.

Left–right differences in the perception of verbal and nonverbal stimuli have been reported in vision (see White, 1969) and audition (e.g., Bryden, 1969; Kimura, 1967). Geffen, Bradshaw, and Wallace (1971), using tachistoscopic presentation of digits and faces, obtained similar differences. Reaction time, rather than a measure such as accuracy of recall, was used to disentangle some possible confounding factors, for example, the effect of memory load and attention on such differences. The same problems arise when auditory inputs are investigated, and for the same reason a time measure, reading time, was used with maximally disruptive delayed auditory feedback (DAF) presented to one ear. Other materials were simultaneously fed to the opposite ear. Hitherto, neither this form of input nor this measure appear to have been employed in a study of left–right effects in audition. Bradshaw (1970) used articulatory lengthening under DAF as a measure of speech disruption.

With dichotic input, verbal material presented to the right ear is more accurately identified and recalled (Broadbent & Gregory, 1964; Bryden, 1963, 1967; Knox & Kimura, 1970; Shankweiler & Studdert-Kennedy, 1966; Studdert-Kennedy & Shankweiler, 1970). Nonverbal but word-like stimuli also show a right-ear superiority (Curry, 1967; Curry & Rutherford, 1967; Haggard, 1969; Kimura, 1967; Kimura & Folb, 1968; Shankweiler & Studdert-Kennedy, 1966, 1967; Zurif & Sait, 1970). Conversely, the left ear has been found superior, with similar dichotic input, in the identification and recall of complex, nonverbal stimuli such as tunes, patterns of tones, etc. (Curry, 1967; Kimura, 1961, 1964, 1967; Knox & Kimura, 1970; Milner, 1962; Spreen, Benton, & Fincham, 1965; Spreen, Spellacy, & Reid, 1970; Vignolo, 1969). The left ear also seems superior in resolving pairs of closely contiguous clicks (Murphy & Venables, 1970). The great majority of studies, including those reported here, have employed right-handed *S*s. This is largely due to the attempt to relate left–right ear differences to cerebral asymmetry, a less complex problem in the case of the right handed. The latter, known as dextrals, are distinguished either by self-classification or the choice of the hand for writing, but the same is not true of left-handed (sinistral) *S*s (Satz, Achenbach, & Fennell, 1967). Of the general population between 5% and 15% have been classified by different

[1] This research was supported by an award from the Australian Research Grants Committee to the senior author. We also wish to acknowledge the assistance given by Diane Nettleton, particularly in connection with the piano-playing tasks.

[2] Requests for reprints should be sent to John L. Bradshaw, Department of Psychology, Monash University, Clayton, Victoria, 3168, Australia.

authorities as "pure," and many more apparent sinistrals, it is claimed, should be reclassified as mixed or ambidextrous (Annett, 1970a, 1970b; Hecaen & Ajuriaguerra, 1964; Kimura & Vanderwolf, 1970). Sinistrals should be distinguished into two groups, familial and nonfamilial (Bryden, 1970; Zurif & Bryden, 1969), depending upon the presence of sinistrality in one or more parents or siblings. Less than half of all sinistrals are familial. Nonfamilial sinistrals perform exactly like dextrals, in terms of ear, visual field and cerebral dominance, while familial sinistrals show much less differentiation and generally greater lateral equipotentiality (Annett, 1970a, 1970b; Branch, Milner, & Rasmussen, 1964; Satz et al., 1967; Zurif & Bryden, 1969). The employment of two groups of *S*s, right- and left-handed, whether familial or otherwise, would thus not resolve the question of whether an ear preference or an underlying mechanism such as cerebral asymmetry determined the effects reported here. However, if a group of right-handed *S*s were shown to be more affected by DAF, in their right ears during a speech task, and in their left ears during a music task, then simple ear preference is unlikely to account for this functional dual asymmetry.

Competitive input has usually been found necessary to produce left–right differences in audition. Several studies ask whether division of attention should be separated from competition per se, (Oxbury, Oxbury, & Gardner, 1967; Satz, 1968; Treisman & Geffen, 1968). Bryden (1969), investigating the relative contributions of expectancy (*S* attends to both ears, not knowing which will be stimulated) and binaural competition, found the latter is necessary and sufficient for producing laterality differences. There is also evidence for asymmetry under certain circumstances between the ears with monaural input (Bakker, 1967a, 1967b, 1968, 1970; Simon, 1967), though not all studies have been successful (Calearo & Antonelli, 1963; Kimura, 1961, 1967). In the visual modality, however, the majority of laterality studies have not employed simultaneous competitive input, either to the two eyes or the two visual hemifields (some exceptions are Dimond, 1969, 1970; Sampson, 1969; Whelan, 1967). Differences in the physiology of the two receptor systems can account for this. In vision, as compared with audition, there is much less neural interaction between the two hemifields before reaching the cortical projection areas. Auditory stimuli arriving at either ear are relayed to both hemispheres (Galambos, 1950).

There is evidence that the inputs on contralateral pathways tend to occlude those arriving simultaneously along the ipsilateral routes before reaching the cortex (Bocca, Calearo, Cassinari, & Migliavacca, 1955; Hall & Goldstein, 1968; Kimura, 1961, 1967; Milner, Taylor, & Sperry, 1968; Sparks & Geschwind, 1968). Rosenzweig (1951) found that in cats, auditory inputs affect the two hemispheres unequally, the amplitude of the ipsilateral response being appreciably less than the contralateral one, though latency and duration were equal. The behavioral studies of Murphy and Venables (1970) tend to confirm such equivalence in transmission times; they obtained a left-ear superiority in the discrimination of closely contiguous clicks only in a choice, never in a simple RT setting. On the other hand, more recent neurophysiological evidence indicates faster contralateral passage in cats (Gross, Small, & Thompson, 1967) and in man (Majkowski, Bochenek, Bochenek, Knapik-Fijalkowska, & Kopeć, 1971). The latter group found a difference of 8 msec.

Four experiments are reported employing unilateral presentation of disruptive DAF, with various other inputs to the opposite ear. The *S*s' reading and piano-playing times were measured. Two questions were asked. Would this technique be sensitive to left–right ear differences? Is competition a necessary precondition, and what type of material is competitive? Finally the nature of attentional processes was considered.

Experiment I

In order to establish whether speech is more affected by disruptive DAF to one

ear or the other, a situation was devised employing the simultaneous presentation to the opposite ear of a second competing input. To make the latter as compatible as possible with the disruptive DAF and *S*'s ongoing speech activity, nondisruptive feedback was employed of a rather longer delay than the other. Two types of reading material were presented, connected prose and lists of common trisyllabic words. The latter all conformed to the same pattern of length and rhythm or stress. It was thought that these, if read at a uniform rate, might approximate somewhat more closely to the paradigm of split-span paired stimulation. This provides for the temporal alignment and simultaneous presentation to each ear of paired items of similar length. It is the paradigm traditionally employed in the recall studies discussed above concerning left–right ear differences. Specifically, it was predicted that reading times would be longer with the disruptive DAF to the right ear, either because it would be more difficult to block out or ignore that channel, or because it would be less easy to switch attention to the less disruptive long-delay feedback in the left ear. Yates (1963) reported that the speech disruption caused by DAF varies for different types of material. It was therefore asked what the effect of the regular, rhythmic word lists might be on any left–right differences.

Method

Apparatus.—Two Revox tape recorders were employed, Model A77 and Model G36. The former was set to give a delay of .2 sec., a value close to that considered maximally disruptive (Yates, 1963). The latter was employed for the long delay (1.1 sec.) since it was easier to modify it by using a 10-cm. tape loop around a roller introduced between the record and playback heads. Other minor modifications were made to help the tape run smoothly. The output from each machine was fed to one of a pair of Maico Auraldomes with TDH 39 headphones and the controls adjusted, using a Brüel and Kjaer precision sound level meter, Type 2203, to give approximate equality at the ear. This resulted in peaks of 90–100 db., on the average, when *Ss* were reading the material under experimental conditions. It was hoped in this way to reduce bone- and stray air-conducted feedback to a minimum, despite some loud articulations. With speech, however, it is impossible to eliminate completely all immediate feedback.

Reading times were obtained with a stop watch and, where necessary, checked from the recording on a third machine (Tandberg 1200 X). This recording was also used to correct articulation times for words omitted, coughs, and the one ocasion when an *S* completely lost her place. The *S* sat at a table with his chest against a padded bar, in front of which the three microphones were arranged. The reading material was placed at eye level immediately behind the microphones.

Material.—The prose passage consisted of Bartlett's (1932) 340-word passage entitled "The war of the ghosts." The word list consisted of 100 items, each three syllables long and of the stress structure exhibited by such words as *signify*, *subsequent*, *stimulate*, etc. Words of doubtful or unfamiliar meaning or pronunciation were excluded.

Procedure.—Sixteen paid student *Ss* were employed, 8 males and 8 females. None had speech defects. All were normally right-handed for all the everyday tasks of life, and without any left-handed close relatives or siblings. These facts were elicited by prior questioning. All *Ss* were given two practice trials on both types of material, once with disruptive DAF to one ear and once to the other. This was achieved simply by reversing the headphones on *S*'s head. Each *S* was tested eight times, twice under each of the four possible experimental conditions (ear by type of material). An ascending–descending counterbalanced design was employed, giving rise to four different orders, within which either the type of material or the ear of presentation was alternated. The *Ss* were instructed to read the material aloud as fast as possible without making mistakes and without correcting themselves.

Results

Average reading times were longer when .2-sec. DAF was played to the right ear rather than the left. Three *Ss* in each case, however, produced contrary results, one of them with both types of material. With the prose passage the average increase was 5.6%; with the word lists the increase was 2.6%. A three-way analysis of variance showed that the laterality effect was significant, F (1, 15) = 9.1, $p < .01$. Despite the reduced magnitude of the effect with the word list, the Laterality × Material interaction failed to reach significance, F (1, 15) = 3.1, $p > .05$. While *Ss* disagreed, on being questioned after the experiment, as to whether they found the long or short delay the more disruptive, the interference does seem to have been of a different nature. The short delay, as ex-

pected, disrupted articulation, and such disturbances were then made apparent to *S* by the long delay. Consequently, while the employment of the long delay might have helped to ensure close similarity between the two kinds of input, it may itself have been disruptive. The *S*s however, were largely unanimous that the disruptive .2-sec. DAF to the right ear was worse than when played to the left. A notable exception was the one *S* whose reading times for both types of material were longer with .2-sec. DAF in his left ear. He could therefore be a case of right-hemsiphere dominance with respect to speech.

Experiment II

The first experiment demonstrates the possibility of producing left–right differences in audition, using DAF and reading times. It can now be asked whether instantaneous auditory feedback (IAF) to the other ear is a sufficiently competitive form of input to restrict the disruptive DAF to its contralateral pathway and demonstrate such differences. The instantenous feedback normally present during speech may be inadequate as a competitive input, *S* perhaps having long habituated to the sound of himself speaking. Experiment II was designed to examine this problem. The IAF was amplified, but only to a level equivalent to that of the DAF.

Method

Apparatus and procedure.—The same two Revox tape recorders were employed. One provided .2-sec. DAF; the other acting as an amplifier produced amplified IAF. The same headphones, sound level meter, and general settings were used as in Exp. I. Only the prose passage was given. Sixteen paid student *S*s were employed, 10 males and 6 females. The instructions and general procedure were the same as in Exp. I. The *S*s were given prior practice and there was a counterbalanced ABBA or BAAB type of design, alternating between *S*s.

Results

The findings were largely similar to Exp. I; there was a 4.7% increase on reading times when the disruptive DAF went to the right ear, and the IAF to the left, as compared with the reverse case. When subjected to a t test for correlated samples, this proved significant; $t(15) = 3.2$, $p < .01$. As with the first experiment, there were three reversals out of 16 *S*s. These results suggest that left–right differences may still occur when *S* hears his own normal air-conducted and bone-conducted feedback through one ear, with a single headphone presenting .2-sec. DAF monaurally to the other ear.

Experiment III

Experiment II demonstrated that the ear differences persist when IAF is used as the alternative input. If competitive input is necessary to produce the effect, then IAF must be adequate as such. Experiment III aimed at determining whether ear differences could still be obtained with white noise presented to the opposite ear. With *S*s of normal hearing, this is, in a sense, the nearest approximation possible to a purely monaural DAF. The white noise should mask as far as possible any stray instantaneous feedback to that ear, while being uninformative, nondistracting, and continuous. If this were to result in some degree of neural habituation, it might even be considered noncompetitive. Consequently, the absence of ear differences under these conditions would tend to support the hypothesis that competitive alternative input is a necessary precondition.

Method

Apparatus and procedure.—White noise was played through a Revox tape recorder, Model A77, from a Brüel and Kjær random noise generator, Type 1402, at a range of 20–20,000 Hz., linear. This substituted for the IAF in Exp. II. The output at the ear was set to 90 db., using the same procedure as before. Again only the prose passage was presented to the 16 paid *S*s, 11 males and 5 females. The same general procedures and settings were employed.

Results

There was a very much smaller percentage increase (1.2%) when .2-sec. DAF was presented to the right ear, as compared with the left. It failed to reach significance,

$t(15) = .06$, $p > .05$, using a t test for correlated samples. While 12 out of the 16 *S*s still showed a performance decrement in the same direction as the previous two experiments, the magnitude of their effects was reduced, and three of the four reversals were large. While these negative results should perhaps be treated with some caution, possibly being due to sampling error, they do, nevertheless, suggest that a simultaneous input of white noise is inadequate in inducing significant ear differences. The slight tendencies noted in the same direction as in the other two experiments could possibly be due to the impossibility of completely masking stray instantaneous feedback in the ear opposite to that receiving the disruptive DAF. In this sense, Exp. III could possibly be considered as a weaker version of Exp. II. It should, however, be noted that *S*s were very obviously still being affected by the DAF; the level of the white noise was not sufficient to mask the input to the other ear. The left–right differences were merely reduced and more variable.

Experiment IV

The first three experiments demonstrated that for right-handed *S*s, the right ear is more affected by disruptive DAF than the left, at least under appropriate conditions of competitive stimulation. Two different explanations seem possible: one based on ear preference and one involving some underlying mechanism such as cerebral asymmetry. If it can be shown that for a given *S*, DAF causes a right-ear decrement in the performance of a speech task, but has the opposite effect on a musical task, then some aspect of hemispheric asymmetry is implicated. Experiment IV was designed to examine this question. It was also asked whether left–right differences could be obtained under conditions involving amplified DAF with attenuated air-conducted IAF to the opposite ear.

Method

Apparatus.—One Revox tape recorder, Model A77, was employed to give a delay of .2 sec. This was close to that considered maximally disruptive for speech (Yates, 1963) and was found, in a pilot experiment, to be more disruptive for piano playing than delays of .4, .55, .75 or 1.1 sec. The music was played on a Danemann upright piano with the microphone placed on the lid. The prose passage was also read at the piano with the script and the microphone both placed on the music stand. The output from the tape recorder was fed to one of a pair of Maico Auraldomes with TDH39 headphones, as in the earlier experiments. The controls were adjusted so that for normal playing or reading, peaks of about 90–100 db. were obtained on a Brüel and Kjaer precision sound level meter, Type 2203. There was no input to the second headphone which thus permitted air-conducted IAF, though attenuated by 20 db. at 2,000 Hz. and 38 db. at 8,000 Hz., to reach one ear.

Material.—The same prose passage as in previous experiments was used. The *S*s were allowed to choose their own music provided it had a fairly fast tempo and could be played at different rates. The music chosen included such varied pieces as "Chopsticks" and a Bach Minuet. Passages taking about 60 sec. to play were chosen. Four *S*s played from a script and four from memory or by ear.

Procedure.—Eight paid student *S*s, 2 males and 6 females were selected on criteria similar to those in the previous experiments. In addition, all could play the piano reasonably well. They were given two practice trials on both tasks, once with the disruptive DAF to one ear, and once to the other. The same counterbalanced design was employed as in Exp. I with the music task replacing the word list. The *S*s were instructed to read the material aloud, or play their piece, as fast as possible without making mistakes and without correcting themselves. For the piano-playing task they were also asked to try and retain the normal rhythm of the piece.

Results

All eight *S*s demonstrated longer reading times with the disruptive DAF played to the right ear (5.4%) and longer playing times with the DAF to the left ear (6.1%). There were no reversals, and the effects for the music task ranged from 2.5% to 13.8%, and for the speech task from 2.1% to 11.0%. The effect of the ear task interaction was significant, $F\ (1, 7) = 52.1$, $p < .001$. It can therefore be concluded that the DAF technique is sensitive to left–right ear differences, and that the direction of these depends upon the nature of the task, and that attenuated IAF will suffice as a competitive input.

Discussion

These four experiments examined the effects of DAF, under conditions of dichotic input,

with right-handed *S*s. They were more affected by the disruptive DAF when it was presented to the right ear in a speech task and to the left ear in a piano-playing task. In the speech tasks, the competing input to the opposite ear varied from long-delayed feedback, amplified IAF, and attenuated IAF, to constant-level white noise. Only in the last case did the left–right differences fail to reach significance. The fact that the direction of the ear preference was shown to depend upon the nature of the task, and that this was in accord with the known phenomena of cerebral asymmetry, would seem strong support for an explanation involving hemispheric differences. Indeed this technique could perhaps provide a convenient and simple tool for determining a patient's dominant hemisphere, avoiding the somewhat hazardous method of injecting sodium amytal into the carotid artery (Wada & Rasmussen, 1960). However in view of the findings relating to familial sinistrals which were summarized earlier, the frequency of actual cases of reversed dominance seems questionable. It is possible that the magnitude of the left–right percentage differences in the two tasks might, nevertheless, reflect the extent to which one or other cerebral hemisphere is preponderant for a particular task. This would be in accord with the claim of Sperry and Gazzaniga (1967) that both hemispheres may subserve some speech functions, though the differences are more manifest when vocal expression is involved.

The findings relating to the requisite nature of the contralateral input can be explained either in terms of a model involving suppression of ipsilateral input or in terms of differential attentional tendencies. The contralateral-suppression model assumes the suppression, partially or completely, of the ipsilateral pathway by the contralateral. In view of the comparatively small though significant ear differences obtained and Rosenzweig's (1951) findings, the ipsilateral suppression would perhaps appear partial. Had temporal alignment of the signals on the two channels been achieved in the word-list condition of the first experiment there might have been more complete ipsilateral suppression. Kimura (1961) and Studdert-Kennedy and Shankweiler (1970) seem to require complete ipsilateral suppression, with left-ear input to the left cerebral hemisphere occurring via the right hemisphere and the corpus callosum. They assume the informational losses to occur during this transfer. The same interpretation is demanded by Geschwind (1970) and Milner et al. (1968). They report the case of a callosal-sectioned patient who processed monaural left-ear input perfectly, but could identify almost nothing when there was competitive right-ear stimulation. Bryden and Zurif (1970) reviewed evidence to suggest that such patients may subsequently recover from their left-ear inabilities under dichotic stimulation. They invoke an explanation similar to the attentional model to be considered below. Possible support for this comes from Sparks and Geschwind (1968), who show that the weakening of the left-ear response depends upon the nature of the sounds in the other ear. As the sounds in the right ear are gradually distorted from speech, the performance on the left ear progressively improves. This could explain the absence of a significant ear effect with contralateral white noise in Exp. III.

The attentional model assumes that speech is largely or entirely located in the left hemisphere and that the latter still maintains its phylogenetically ancient role of monitoring events to the right of the body. Consequently, when *S* attends to a speech task, he finds it easier to attend to the right ear (cf. Kinsbourne, 1970). The *S*s spontaneously commented in the experiments reported here that it was more difficult to ignore the disruptive DAF to the right ear for the speech tasks, and to the left ear for music. The attentional model would suggest that left–right differences would be a function of the nature and requirements of the overall task. Darwin (1971) cites evidence to show that for dichotic presentation of vowels, the right-ear advantage occurs only in the presence of acoustic and *phonemic* uncertainty and that a left-ear advantage may occur for the same stimuli in a nonspeech, musical context. For a speech context, perhaps *acoustic* features can be correctly extracted in either hemisphere, but can only be related to *phonemic* features in a verbal context in the left hemisphere (cf. Studdert-Kennedy & Shankweiler, 1970).

The attentional model would predict a greater ear difference in Exp. II than in Exp. I since the IAF should have been more easily attended to than the 1.2-sec. DAF. This however was not the case. Neither model predicts a left–right difference in Exp. III if the constant white noise neither provides a satisfactory alternative channel to attend to nor is adequate in suppressing the ipsilateral route. Two studies reported since the completion of the present experiments sought to determine

whether, in a shadowing task, *S*s could voluntarily ignore the effects of disruptive DAF introduced into the nonattended channel. Robinson (1970) found that they could not, while Zelniker (1970) obtained the opposite result.

In conclusion, it is felt that an explanation involving both cerebral asymmetry and at least some aspects of attention is intuitively more satisfying, and seems to accord better with the known facts. Speech and piano playing always involve some air-conducted, if not bone-conducted feedback; therefore, it will not be possible to determine whether left–right differences are obtainable in a purely monaural, noncompetitive situation, unless some task is employed where the sound originates and is modified electronically.

REFERENCES

ANNETT, M. A classification of hand preference by association analysis. *British Journal of Psychology*, 1970, **61**, 303–321.(a)

ANNETT, M. The growth of manual preference and speed. *British Journal of Psychology*, 1970, **61**, 545–558.(b)

BAKKER, D. J. Ear asymmetry with monaural stimulation; Task influence. *Cortex*. 1967, **5**, 36–42. (a).

BAKKER, D. J. Left-right differences in auditory perception of verbal and nonverbal material by children. *Quarterly Journal of Experimental Psychology*, 1967, **19**, 334–336. (b).

BAKKER, D. J. Ear asymmetry with monaural stimulation. *Psychonomic Science*, 1968, **12**, 62.

BAKKER, D. J. Ear asymmetry with monaural stimulation: Relations to lateral dominance and lateral awareness. *Neuropsychologia*, 1970, **8**, 103–117.

BARTLETT, F. C. *Remembering*. London: Cambridge University Press, 1932.

BOCCA, E., CALEARO, C., CASSINARI, V., & MIGLIAVACCA, F. Testing "cortical" hearing in temporal lobe tumours. *Acta Oto-laryngologica*, 1955, **45**, 289–304.

BRADSHAW, J. L. Phonetic homogeneity and articulatory lengthening. *British Journal of Psychology*, 1970, **61**, 499–507.

BRANCH, C., MILNER, B., & RASMUSSEN, T. Intracarotid sodium amytal for the lateralization of cerebral speech dominance. *Journal of Neurosurgery*, 1964, **21**, 399–405.

BROADBENT, D. E., & GREGORY, M. Accuracy of recognition for speech presented to right and left ears. *Quarterly Journal of Experimental Psychology*, 1964, **16**, 359–366.

BRYDEN, M. P. Ear preference in auditory perception. *Journal of Experimental Psychology*, 1963, **65**, 103–105.

BRYDEN, M. P. An evaluation of some models of dichotic listening. *Acta Oto-laryngologica*, 1967, **63**, 595–604.

BRYDEN, M. P. Binaural competition and division of attention as determinants of the laterality effect in dichotic listening. *Canadian Journal of Psychology*, 1969, **23**, 101–113.

BRYDEN, M. P. Laterality effects in dichotic listening: Relations with handedness and reading ability in children. *Neuropsychologia*, 1970, **8**, 443–450.

BRYDEN, M. P., & ZURIF, E. B. Dichotic listening performance in a case of agenesis of the corpus callosum. *Neuropsychologia*, 1970, **8**, 371–377.

CALEARO, C., & ANTONELLI, A. R. "Cortical" hearing tests and cerebral dominance. *Acta Oto-laryngologica*, 1963, **56**, 17–26.

CURRY, F. K. W. A comparison of left- and right-handed subjects on verbal and nonverbal dichotic listening tasks. *Cortex*, 1967, **3**, 343–352.

CURRY, F. K. W., & RUTHERFORD, D. R. Recognition and recall of dichotically-presented verbal stimuli by right- and left-handed persons. *Neuropsychologia*, 1967, **5**, 119–126.

DARWIN, C. J. Ear differences in the recall of fricatives and vowels. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 46–62.

DIMOND, S. J. Hemisphere function and immediate memory. *Psychonomic Science*, 1969, **16**, 111–112.

DIMOND, S. J., Hemispheric refractoriness and control of reaction time. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 610–617.

GALAMBOS, R. Neurophysiology of the auditory system. *Journal of the Acoustical Society of America*, 1950, **22**, 785–791.

GEFFEN, G., BRADSHAW, J. L., & WALLACE, G. Interhemispheric effects on reaction time to verbal and nonverbal visual stimuli. *Journal of Experimental Psychology*, 1971, **87**, 415–422.

GESCHWIND, N. The organization of language and the brain. *Science*, 1970, **170**, 940–944.

GROSS, N., SMALL, A., & THOMPSON, D. Response to contralateral and ipsilateral auditory stimulation from the same cortical areas. *Brain Research*, 1967, **5**, 250–262.

HAGGARD, M. P. Perception of semivowels and laterals. *Journal of the Acoustical Society of America*, 1969, **46**, 115.

HALL, J. L., & GOLDSTEIN, M. H. Representation of binaural stimuli by single units in primary auditory cortex of unanaesthetized cats. *Journal of the Acoustical Society of America*, 1968, **43**, 456–461.

HECAEN, H., & AJURIAGUERRA, J. DE. *Left handedness*. New York: Grune & Stratton, 1964.

KIMURA, D. Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology*, 1961, **15**, 166–171.

KIMURA, D. Left-right differences in the perception of melodies. *Quarterly Journal of Experimental Psychology*, 1964, **16**, 355–358.

KIMURA, D. Functional asymmetry of the brain in dichotic listening. *Cortex*, 1967, **3**, 163–178.

KIMURA, D., & FOLB, S. Neural processing of backwards speech sounds. *Science*, 1968, **161**. 395–396.

KIMURA, D., & VANDERWOLF, C. H. The relation between hand preference and the performance of individual finger movements by left and right hands. *Brain*, 1970, **93**, 769–774.

KINSBOURNE, M. The cerebral basis of lateral asymmetries in attention. *Acta Psychologia*, 1970, **33**, 193–201.

KNOX, C., & KIMURA, D. Cerebral processing of nonverbal sounds in boys and girls. *Neuropsychologia*, 1970, **8**, 227–237.

MAJKOWSKI, J., BOCHENEK, Z., BOCHENEK, W., KNAPIK-FIJALKOWSKA, D., & KOPEĆ, J. Latency of averaged evoked potentials to contralateral and ipsilateral auditory stimulation in normal subjects. *Brain Research*, 1971, **25**, 416–419.

MILNER, B. Laterality effects in audition. In V. B. Mountcastle (Ed.), *Interhemispheric effects and cerebral dominance.* Baltimore: Johns Hopkins Press, 1962.

MILNER, B., TAYLOR, L., & SPERRY, R. W. Lateralized suppression of dichotically presented digits after commissural section in man. *Science*, 1968, **161**, 184–185.

MURPHY, E. H., & VENABLES, P. H. The investigation of ear asymmetry by simple and disjunctive reaction-time tasks. *Perception & Psychophysics*, 1970, **8**, 104–106.

OXBURY, S. M., OXBURY, J. M., & GARDINER, J. Laterality effects in dichotic listening. *Nature*, 1967, **214**, 742.

ROBINSON, D. O. The limits of selective attention in DAF shadowing. *Psychonomic Science*, 1970, **21**, 325–327.

ROSENZWEIG, M. R. Representations of the two ears at the auditory cortex. *American Journal of Physiology*, 1951, **167**, 147–158.

SAMPSON, H. Recall of digits projecting to temporal and nasal hemiretinas. *Quarterly Journal of Experimental Psychology*, 1969, **21**, 39–42.

SATZ, P. Laterality effects in dichotic listening. *Nature*, 1968, **218**, 277–278.

SATZ, P., ACHENBACH, K., & FENNELL, E. Correlations between assessed manual laterality and predicted speech laterality in a normal population. *Neuropsychologia*, 1967, **5**, 295–310.

SHANKWEILER, D., & STUDDERT-KENNEDY, M. Lateral differences in perception of dichotically-presented synthetic consonant-vowel syllables and steady-stage vowels. *Journal of the Acoustical Society of America*, 1966, **39**, 1256.

SHANKWEILER, D., & STUDDERT-KENNEDY, M. Identification of consonants and vowels presented to left and right ears. *Quarterly Journal of Experimental Psychology*, 1967, **19**, 59–63.

SIMON, J. R. Ear preference in a simple reaction time task. *Journal of Experimental Psychology*, 1967, **75**, 49–53.

SPARKS, R., & GESCHWIND, N. Dichotic listening in man after section of neocortical commissures. *Cortex*, 1968, **4**, 3–16.

SPERRY, R. W., & GAZZANIGA, M. S. Language following surgical disconnection of the hemispheres. In C. H. Millikan & F. L. Darley (Eds.), *Brain mechanisms underlying speech and language.* New York: Grune & Stratton, 1967,

SPREEN, O., BENTON, A. L., & FINCHAM, R. W. Auditory agnosia without aphasia. *Archives of Neurology*, 1965, **13**, 84–92.

SPREEN, O., SPELLACY, F. J., & REID, J. R. The effect of interstimulus interval and intensity on ear asymmetry for nonverbal stimuli in dichotic listening. *Neuropsychologia*, 1970, **8**, 245–250.

STUDDERT-KENNEDY, M., & SHANKWEILER, D. Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, 1970, **48**, 579–594.

TREISMAN, A. M., & GEFFEN, G. Selective attention and cerebral dominance in perceiving and responding to speech messages. *Quarterly Journal of Experimental Psychology*, 1968, **20**, 139–150.

VIGNOLO, L. A. Auditory agnosia: A review and report of recent evidence. In A. L. Benton (Ed.), *Contributions to clinical neuropsychology*, Chicago: Aldine, 1969.

WADA, J., & RASMUSSEN, T. Intracarotid injection of sodium amytal for the lateralization of cerebral speech dominance: Experimental and clinical observations. *Journal of Neurosurgery*, 1960, **17**, 266–282.

WHELAN, E. Visual perception and cerebral dominance. Unpublished doctoral dissertation, Sheffield University, 1967.

WHITE, M. J. Laterality differences in perception: A review. *Psychological Bulletin*, 1969, **72**, 387–405.

YATES, A. J. Delayed auditory feedback. *Psychological Bulletin*. 1963, **60**, 213–232.

ZELNIKER, T. Perceptual attenuation of an irrelevant auditory verbal input as measured by an involuntary verbal response in a selective attention task. *Journal of Experimental Psychology*, 1970 **87**, 52–56.

ZURIF, E. B., & BRYDEN, M. P. Familial handedness and left-right differences in auditory and visual perception. *Neuropsychologia*, 1969, **7**, 179–187.

ZURIF, E. B., & SAIT, P. E. The role of syntax in dichotic listening. *Neuropsychologia*, 1970, **8**, 239–244.

(Received April 11, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 93-97

ARITHMETIC PROBLEM SOLVING[1]

MARGARET JEAN PETERSON AND SONIA ALLER[2]

Indiana University

The experiment measured latencies as college students solved problems involving one or more arithmetic operations. Simple addition and multiplication problems were solved more quickly than negative addition or subtraction problems. All of the problems testing multiple arithmetic operations required more time to solve than the problems testing a single operation. The problems also differed in the numerical values assumed by the larger number given in the problem, by the smaller number given in the problem, and by the unknown. These variables were complexly related to the solution times.

Most educated adults can solve, without the use of paper and pencil, equations involving one or two arithmetic operations on relatively small numbers, but little research appears to have been done on this kind of cognitive skill. For example, the following four equations use similar numbers, but the arithmetic operations are assumed to be addition, subtraction, multiplication, and "negative addition," respectively: $3 + x = 8$; $8 - x = 3$; $3(x) = 15$; $8 + x = 3$. Negative addition problems require subtraction for their solution, but only positive values are indicated in the equation. One purpose of this study was to compare the solution latencies of these four types of problems.

Another purpose was to examine the time required to solve problems that combined two or more of the simple arithmetic operations. One combination of two single-operation problems such as $3 + x = 8$ and $8(x) = 40$ could be the multioperation problem, $8(3 + x) = 64$. If the multioperation problem represents a simple summation of the solution processes for the two single-operation problems, then the latency required to solve the multioperation problem would equal the sum of the latencies required to solve the single-operation problems. This prediction tenuously assumes that the time to read in, solve, and report out the solutions to the two single-operation problems approximates the time to read in, solve, and report out the solution of the multioperation problem.

The solution times of these types of problems also may be influenced by the numerical values stated in the problems. Moyer and Landauer (1967) and Restle (1970) demonstrated that solution latencies were affected by the difference between quantities stated in the problems. Moyer and Landauer's *S*s identified the larger of two single digits, a or b, and Restle's *S*s indicated the larger of two quantities, $(a + b)$ or c. In both experiments, the latencies tended to increase as the difference between a and b decreased, except, in the Restle situation, when $a = b$. When $a = b$, the latencies were as short as with maximal differences between a and b. These studies used college students as *S*s.

When elementary school children solved simple addition, subtraction, and multiplication problems, Suppes, Hyman, and Jerman (1967) reported that magnitude variables such as the value of the sum and of the smallest addend in addition problems did not predict performance as satisfactorily as the total number of steps assumed to change the problem to canonical form and to perform operational and memorial steps. The discrepant results may have been produced by differences in the tasks, *S*s, or the particular magnitude relationships used. However, the extent of the effect of magnitude variables upon adult performance seems ambiguous and the interaction with various arithmetic operations has not been explored. Consequently,

[1] The authors wish to express appreciation to F. Restle and L. Peterson for their very helpful discussions.

[2] Requests for reprints should be sent to M. J. Peterson, Department of Psychology, Indiana University, Bloomington, Indiana 47401.

the magnitudes assumed by the numbers given in both single-operation and multi-operation problems were varied in the present experiment.

Lastly, the position of the unknown varied because Groen (1967) and Suppes et al. (1967) found that the latencies were longer when the unknown was in the initial position, $x + 3 = 8$, than when it was embedded, $3 + x = 8$.

Method

Subjects.—Sixty-three students in introductory psychology courses at Indiana University participated to fulfill course requirements. They were assigned randomly to one of three starting positions in the list of materials. Twenty-four students made errors on 20 problems or more. They were judged to be noncompetent arithmetically and their data were discarded. The remaining 39 *S*s contributed the results presented below.

Materials.—Each *S* was shown six sets of 20 problems. The six sets differed in the values assigned to the three magnitude variables: the unknown (x), the smaller number (Small), and the larger number (Large). One value was determined for each of the magnitude variables to be used in all problems of one set. The number to the right of the equal sign varied within the sets to satisfy the required arithmetic operations. The values used for the six sets were, in order of x, Small, and Large: Set A = 2, 6, 8; Set B = 3, 2, 5; Set C = 4, 3, 7; Set D = 5, 3, 8; Set E = 6, 5, 11; Set F = 8, 1, 9.

Each set contained 20 problems and the problems followed identical formats for each set. As shown in Table 1, eight problems compared simple addition, negative addition, subtraction, and multiplication operations. The two problems testing each operation differed only in the placement of the unknown. Thus, for the addition problems, x was embedded (Problem 1) or was in the initial position (Problem 2). Problems 3 and 4 tested negative addition; Problems 5 and 6 tested subtraction; Problems 7 and 8, multiplication.

The last 12 problems dealt with combinations of arithmetic operations. Problems 9–12 compared the effect of multiplication of a quantity that indicated either addition or negative addition. The unknown occupied the first or second position within the parentheses of the quantity. Problems 13–16 presented multiplication of x by a quantity involving either addition or subtraction of Large and Small. The positions of Large and Small within the parentheses varied. The last four problems were concerned with the multiplication of x by Large plus the addition of Small or the multiplication of x by Small and the addition of Large. The position of x varied.

The 120 equations were randomized and typed in three columns of 40 equations on a memory drum tape. The latencies were measured by a Standard Electric timer. The timer was activated simultaneously with a single pulsing of the memory drum by *E*'s depression of a push button. The timer was stopped by *S*'s pressing his push button.

Procedure.—All *S*s were shown three practice equations printed on 3 × 5 in. cards, and then the 120 equations were displayed on the memory drum tape. Each equation was shown until *S* pressed his push button to indicate he knew the value of x. He spoke his answer aloud. Accuracy and speed were stressed in the instructions. No corrections or feedback were given.

Scoring.—The mean latencies for the correct responses (success latencies) are given in Table 1. Each entry was based on scores from 27 or more *S*s. The errors tended to be associated with the problems that had longer success latencies; hence, the means in Table 1 present a conservative estimate of the differences. Most errors occurred on negative addition problems because *S* did not state the minus sign. The entries do not include deviant latencies. Deviant latencies were defined as success latencies that differed from the next closest latency by 3 sec. For example, the longest success latency on Problem 20, Set A, 11.87, was excluded because it exceeded the next closest latency, 7.28, by more than 3 sec. A similar procedure was followed for very short latencies. It was assumed that such deviant latencies arose from failures to attend immediately or from anticipatory responding by *S*.

Results

Success latencies, Problems 1 to 8.—Problems 1–8 compared the speed of solving simple addition, negative addition, subtraction, and multiplication problems as a function of the type of operation, the six number sets, and the position of the unknown. Instead of doing a least-squares analysis of variance, missing scores were filled by the means for each problem of each set. This procedure was followed to facilitate subsequent analyses of the additivity of the single-operation problems.

The mean success latencies for addition (2.28) and multiplication (2.33) were reliably faster than those for negative addition (2.86) and subtraction (3.33), F (3, 114) = 82.54 (all ps < .01 unless otherwise indicated). The mean latency for subtraction was significantly slower than the means for any other single-operation problem.

The mean success latencies for the six number sets also differed, F (5, 190) = 13.93, primarily because Set B's problems were solved more rapidly than prob-

TABLE 1

PROBLEM FORMATS, SAMPLE PROBLEMS, AND MEAN SUCCESS LATENCIES

Problem nos.	Problem format	Sample problem (Set B)	Mean success latencies						$\bar{X}$-over no. set
			Set A	Set B	Set C	Set D	Set E	Set F	
1 and 2	Small + x = Large	$2 + x = 5$	2.26	2.34	2.54	2.14	2.21	2.16	2.28
3 and 4[a]	Large + x = Small	$5 + x = 2$	3.02	2.49	2.74	2.78	3.30	2.85	2.86
5 and 6	Large − x = Small	$5 - x = 2$	3.41	2.88	3.06	3.54	3.19	3.92	3.33
7 and 8	Large (x) = a	$5(x) = 15$	2.38	1.96	2.30	2.64	2.07	2.64	2.33
9 and 10	Large (Small + x) = a	$5\ (2 + x) = 25$	3.98	4.22	4.27	4.38	6.97	4.44	4.71
11 and 12[a]	Small (Large + x) = a	$2\ (5 + x) = 4$	4.97	5.79	7.32	5.81	5.56	5.66	5.85
13 and 14	x (Small + Large) = a	$x\ (2 + 5) = 21$	4.52	3.20	3.38	3.72	9.38	3.28	4.58
15 and 16	x (Large − Small) = a	$x\ (5 - 2) = 9$	3.20	4.76	3.72	3.19	4.26	3.63	3.79
17 and 18	Large (x) + Small = a	$5\ (x) + 2 = 17$	5.57	3.76	5.57	3.80	5.94	5.02	4.94
19 and 20	Small (x) + Large = a	$2\ (x) + 5 = 11$	4.54	3.96	5.56	6.12	5.88	6.16	5.37
			$\bar{X}$ 3.78	3.54	4.05	3.81	4.88	3.98	

Note.—For Set A, $X = 2$, Large = 8, and Small = 6; for Set B, $x = 3$, Large = 5, and Small = 2; for Set C, $x = 4$, Large = 7, and Small = 3; for Set D, $x = 5$, Large = 8, and Small = 3; for Set E, $x = 6$, Large = 11, and Small = 5; for Set F, $x = 8$, Large = 9, and Small = 1.

[a] x assumes a negative value.

lems of other number sets. The latencies for the other number sets did not differ reliably.

The type of problem interacted with number sets, $F\ (15, 570) = 8.88$. The mean latencies did not differ for the six sets of addition problems. On multiplication and subtraction problems, the mean latencies increased reliably from Set B to Set F, and for negative addition problems Set B also yielded the lowest mean but Set E produced the highest. These results deviated somewhat from predictions made by the magnitude variables, Small and Large. These variables would predict that the mean success latencies should differ for the six number sets, corresponding to increases in the values assumed by x, Small, or Large or to increases in various relationships between the given values. The increase from Set B to Set E for negative addition problems corresponded to a single magnitude variable, Large, and the increases from Set B to Set F for the multiplication and subtraction problems would have been predicted by a summation of the values of x and Large. No other relationship yielded an expected maximum latency for Set F.

The interaction between type of problem and position of x was significant, $F\ (3,114) = 21.24$, subtraction requiring more time for solution when x was in the initial position ($\bar{X} = 3.72$) than when it was embedded ($\bar{X} = 2.94$). The position of x did not differentially affect the three other types of problems.

Success latencies, Problems 9 to *12*.—The problems requiring addition plus multiplication (9 and 10) were solved more quickly ($\bar{X} = 4.71$) than the problems involving negative addition (11 and 12, $\bar{X} = 5.85$), $F\ (1, 38) = 38.69$. The type of problem interacted with the number sets, $F\ (5, 190) = 21.97$, because Set E produced the longest mean latency for addition and Set C, for negative addition. Set A yielded the shortest mean latency for both types of operations, but the differences between the means for all sets except those producing the longest latencies were not statistically significant.

The magnitude variable, Large, would have predicted the long latency for Set E, but neither Large nor the other magnitude variables indicated that Set C should yield long latencies for any of the problems.

The position of x interacted with the problem types, $F\ (1, 38) = 8.05$. Solution speed was faster for addition problems when x was embedded ($\bar{X} = 4.65$) than when x was in the first position within the parentheses ($\bar{X} = 4.77$), but the opposite was true for the negative addition problems. These problems were solved faster

when x was in the initial position (5.62) than in the second position (6.08) within the parentheses.

Success latencies, Problems 13 to 16.—The problems requiring addition within the parentheses (13 and 14) yielded longer solution latencies (4.58) than the problems requiring subtraction in the parentheses (15 and 16, $\bar{X} = 3.79$), $F\ (1, 38) = 41.40$. Both types of problems were easier when Large preceded Small within the parentheses, $F\ (1, 38) = 6.15$. The type of problem and the order of presenting Large and Small did not interact.

The number sets also yielded different success latencies, $F\ (5, 190) = 67.03$. Set E had reliably longer success latencies than any other sets.

Success latencies, Problems 17 to 20.—Problems 17 and 18 indicated that x was to be multiplied by Large and added to Small. They had a shorter mean latency (4.94) than Problems 19 and 20, which multiplied x by Small and added Large (5.37), $F\ (1, 38) = 10.86$.

The average success latency for the six sets varied, $F\ (5, 190) = 16.16$, and the interaction with problems was reliable, $F\ (5, 190) = 5.36$. Problems 17 and 18 of Sets B and D required shorter solution times than the same problems of the other sets, and Problems 19 and 20 of Sets B and A were solved significantly faster than the same problems of other sets. The success latencies of the other sets did not differ among each other for either Problems 17 and 18 or Problems 19 and 20, nor did the placement of x produce differential success latencies.

TABLE 2

TESTS FOR ADDITIVITY

Comparison		Success latencies	
Single-Operation problems	Multi-operation problems	Predicted	Observed
1 + 7	9	4.78	4.60
1 + 7	18	4.78	5.01
2 + 7	10	4.58	4.73
2 + 8	17	4.45	4.81
3 + 7	11	5.16	6.00
4 + 7	12	5.35	5.63

Additivity.—A last question asked whether the success latencies required to solve multistep problems could be predicted accurately from the performance on simpler problems that represented component steps of the complex problems. For example, the addition of the steps of Problem 1 to those of Problem 7 yielded the steps assumed to be required in Problem 18. The differences between the latency predicted by the sum of the mean latencies on Problems 1 and 7 and the mean latency obtained on Problem 18 were computed for each S on each number set, and an analysis of variance was applied to the differences. The differences between the latencies predicted from a simple summation of the two parts represented by the latencies on Problems 1 and 7 and the mean latency for Problem 18 did not differ significantly, $F\ (1, 38) < 1$. The same type of comparison was performed for the problems given in Table 2. The value of F was nonsignificant in each case. However, for each of the analyses the differences between the number sets were reliable, indicating that additivity did not hold for all sets.

DISCUSSION

The results indicated that different arithmetic operations require different lengths of time for solution even when similar numbers are presented by the problems. Simple addition and multiplication problems were solved more rapidly by adults than negative addition and subtraction problems. As would be expected, the latencies to solve problems containing multiple arithmetic operations exceeded the latencies required to solve single-operation problems. The multioperation problems with the form x (Large − Small) had the shortest mean success latencies, and the longest latencies were produced by the format Small (Large + x) when negative addition within the parentheses was required for solution.

The data also indicated that in general the sum of the success latencies required for single-operation problems did not adequately predict the latency for multioperation problems. This failure might have occurred because the processing of the problems actually was nonadditive or because the comparisons confounded input–output times with the latencies of the

mental operations involved in solving the problems. The sum of the latencies for the single-operation problems included two sets of input–output times, while the latencies for the multioperation problems contained one. Furthermore, it is possible that *S* occasionally forgot partial solutions and had to repeat his solution process and that idiosyncratic methods for solving combinatorial problems also influenced the solution times.

The six number sets yielded different success latencies, although the ordering of the number sets varied from problem to problem. The Number Sets × Problems interaction indicated that no single magnitude variable or relationship between the variables satisfactorily predicted the latencies, for if such variables were the sole determiners, approximately the same ordering of the number sets would be found for each problem.

Large magnitudes of the component numbers may affect the latencies by overloading the memory of *S*. This notion would be consistent with the frequency of Set E, with the greatest value of Large, to elicit the longest solution times. If it is further assumed that the various problems also depend upon memory, these two characteristics of the equations would have an interactive effect upon the solution times.

The problems could be related to memory in the following way. Suppose that arithmetically competent adults have stored arithmetic "facts" which they may consult as they would any other stored memory. The memory facts might vary in accessibility, necessitating longer searches for less accessible facts, such as those involved in negative addition and subtraction. The longer solution latencies of the multioperation problems might result from the memory searches plus the extra step of holding a partial solution in memory.

The position of x affected performance on only a few problems. Solution was faster when x was embedded in simple subtraction problems and in multioperation problems which included addition in the form Large (Small $+ x$). In the multioperation, negative addition problems, Small ($x +$ Large), solution was faster when x was the first entry inside the parentheses. In general, arithmetic problem solving by adults did not appear to be as sensitive to the placement of the unknown as problem solving by the elementary school children of Groen (1967) and Suppes et al. (1967).

REFERENCES

Groen, G. J. *An investigation of some counting algorithms for simple addition problems.* (Tech. Rep. No. 118) Stanford, Calif.: Stanford University Institute for Mathematical Studies in the Social Sciences, 1967.

Moyer, R. S., & Landauer, T. K. Time required for judgments of numerical inequality. *Nature,* 1967, **215**, 1519–1520.

Restle, F. Speed of adding and comparing numbers. *Journal of Experimental Psychology,* 1970, **83**, 274–278.

Suppes, P., Hyman, L., & Jerman, M. Linear structural models for responses and latency performance in arithmetic. In J. P. Hill (Ed.), *Minnesota Symposia on Child Psychology.* Minneapolis: University of Minnesota Press, 1967.

(Received April 29, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 98–102

POSITIVE TRANSFER BETWEEN NONREWARD AND DELAY

MITRI E. SHANAB [1]

Fresno State College

Following 56 trials of training under either continuous reward, continuous delay of reward, partial reward, or partial delay of reward, all *Ss* were shifted to continuous delay for 28 trials. It was found that throughout the shift phase the partial reward and partial delay groups ran significantly faster than the continuous reward group. The results were interpreted as supporting A. Amsel's new theory of persistence.

Amsel (in press) recently proposed a general theory of persistence that treats the partial reinforcement effect (PRE) as a special case. According to this theory, all anticipatory frustration or punishing events initially interfere with the ongoing goal-approach response, but later become counterconditioned to the goal-approach behavior. Thus frustration produced by partial reward (Amsel, 1958), or partial delay of reward (Donin, Surridge, & Amsel, 1967) and fear produced by partial punishment (Banks, 1966) all result in a PRE reflecting a dimension of behavior which Amsel (1969) identifies as persistence. Persistence is assumed to develop whenever an organism learns through counterconditioning to maintain an ongoing response in the presence of any kind of cues which usually evoke competing or disruptive responses. This implies that persistence effects are not specific to particular stimuli but can transfer to other similar stimuli. Thus Amsel (1969) has proposed that persistence acquired under one disruptive situation would transfer to another disruptive situation. However, the degree of such transfer remains to be determined.

Several lines of evidence suggest that some significant transfer takes place between fear and frustration (cf. Brown & Wagner, 1964). In the Brown and Wagner study, the transfer reported was not complete in that during extinction, the partially reinforced group was significantly faster than the previously punished group shifted to nonreward. However, under conditions of continuous reward and punishment, neither group was significantly different from the other, although graphically the punished group appeared to run consistently faster than the partially reinforced group.

A similar attempt was made by Glazer and Amsel (1970) to test the amount of transfer that takes place between two different frustation-producing operations: partial reward versus partial goal blocking. The procedure used for goal blocking was to cover the food cup with "highly perforated" Plexiglas, making the food inaccessible yet tantalizing. The strong smell of inaccessible food could have acted as a a punishing condition, thus accounting for the lack of complete transfer as was the case in the study of Brown and Wagner (1964). It is conceivable that the lack of complete transfer was due to what Glazer and Amsel called "different but overlapping systems controlling persistence . . . [p. 311]." This implies that if more similar systems are used, more complete transfer would take place.

The present study sought to test the above implication by measuring the degree of transfer between two frustration-producing conditions, viz., partial delay and partial reward, which are assumed to have more similar underlying systems than the previous punishment and goal-blocking conditions. To be sure, delay could be considered punishing; however, to the extent that all trials under either continuous or partial delay are reinforced, the aversiveness of delay might be attenuated relative to a regular punishment procedure using

[1] Requests for reprints should be sent to Mitri E. Shanab, Department of Psychology, Fresno State College, Fresno, California 93710.

electric shock or the "see-but-do-not-touch" goal-blocking procedure.

If the foregoing reasoning is valid, then Ss previously receiving either partial reward or partial delay of reward would run significantly faster than a continuously rewarded group when all Ss are shifted to continuous delay conditions. Moreover, complete transfer would be reflected in a lack of difference between the two partial groups.

METHOD

Design.—The study consisted of two phases. In Phase 1, a 2 × 2 design was used in which different Ss received either continuous reward, continuous delay of reward, partial reward, or partial delay of reward. In Phase 2, all Ss received continuous delay conditions. The continuous delay group was included in the design mainly to provide information about the durability of the transfer effects.

Subjects.—The Ss were 40 male albino rats of the Sprague-Dawley strain, approximately 90 days old at the beginning of the experiment.

Apparatus.—A 1.5-m. runway made of unpainted redwood was used. The runway was covered with Plexiglas and was 23 cm. high throughout. The start box was 18 cm. long and 17 cm. wide, while the goal box was 30.5 cm. long and 15 cm. wide. The runway section was 10 cm. wide. Four sets of photocells were installed in the runway. Interruption of any of the four photobeams could start and/or stop one of three Standard Electric timers which measured start, run, and goal times. The first and second photocells were located 6.4 cm. and 21.5 cm. from the start box, respectively. The third photocell was located 103 cm. from the start box, and the last photocell was located 11.4 cm. within the goal box. Only run times were recorded.

Procedure.—Upon arrival from the supplier, Ss were placed on free feeding for 3 days, following which, and for the duration of the experiment, the rats were placed on a restricted ration of 12 gm. per day. Water, however, was available all the time. During the first week of deprivation, each S was handled and encouraged to eat four 45-mg. food pellets from a 10-cm.-diameter glass cup later used in the goal box. Following this handling period, each S was allowed to explore the runway for 3 min. per day for 3 days. No food was given in the runway during exploration, and all circuitry was turned on to adapt S to the various noises of the equipment. Following the exploration period, Phase 1 was started and lasted for 56 trials. On the first 2 days of Phase 1, each S received two daily trials, each according to its group. Thereafter, each S received four daily trials. The reinforcer consisted of four 45-mg. Noyes pellets. For Ss in the partial delay and partial reward groups, a quasi-random schedule of reinforcement was constructed such that each S received an equal number of rewarded and nonrewarded (or delayed) trials, with the restriction that no more than two rewarded or nonrewarded (or delayed) trials occur in a row. On either nonrewarded or delayed trials, S was confined in the goal box for 15 sec. On nonrewarded trials, an empty glass cup, which was different from the regular food cup used on rewarded trials, was placed in the goal box. On delayed trials, S ran to an empty goal box, and 15 sec. later, the baited cup was lowered into the goal box. For Ss in the continuous reward group, each trial was rewarded. The Ss in the continuous delay group received the same treatment as Ss in the partial delay group on delayed trials. All Ss were removed from the goal box following consumption of the food reinforcer.

In Phase 2, all Ss were delayed 15 sec. on each of their four daily trials. The procedure used was identical to the one used in Phase 1 on delayed trials. In both phases, the intertrial interval was approximately 10 min. Phase 2 was continued until each S had received a total of 28 trials.

RESULTS

All analyses reported here are based on mean run speeds.

Phase 1.—A visual inspection of terminal speeds indicated that the groups were performing at a stable level. This conclusion was supported by an appropriate analysis of variance test with repeated measures over the last three blocks (12 trials). Both the schedule and the delay effects were significant, $F\ (1, 36) = 6.81$, $p < .05$, for schedule, and $F\ (1, 36) = 13.29$, $p < .001$, for delay. The Schedule × Delay interaction was significant, $F\ (1, 36) = 44.55$, $p < .001$. The effect of blocks was not significant, $F\ (2, 72) = 1.34$, $p > .05$, indicating stable performance for all four groups over the last 12 trials in Phase 1. None of the three interactions with blocks was significant. Tests of simple main effects revealed that Ss in the continuous delay group ran significantly slower than Ss in the other three groups, $F\ (1, 36) = 53.25$, $F\ (1, 36) = 19.55$, and $F\ (1, 36) = 43.09$, all $ps < .001$, for the comparisons with the continuous reward, partial reward, and partial delay groups, respectively. No other comparisons were significant.

Phase 2.—As can be seen in Fig. 1, the partial reward and partial delay groups shifted to continuous delay showed only a moderate decrement in mean run speed,

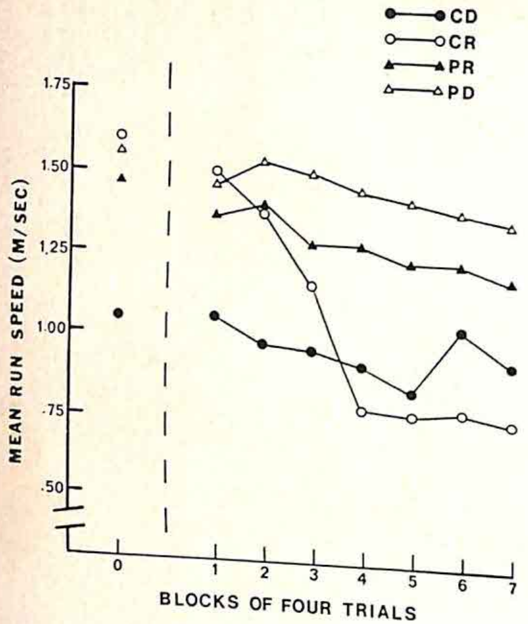

FIG. 1. Mean run speed as a function of a shift from either partial delay (PD), partial reward (PR), or continuous reward (CR) to continuous delay of reward (CD). (Block O denotes preshift asymptotic levels of the four experimental conditions.)

while the continuously rewarded *S*s showed a considerable decrement in performance which apparently undershot the level of the continuous delay control group. The results of the analysis of variance test on the mean run speeds over the entire phase revealed a significant schedule effect, F (1, 36) = 45.16, $p < .001$). The delay factor was not significant, ($F < 1$), while the Schedule × Delay interaction approached significance, F (1, 36) = 4.00, $.05 < p < .10$. The blocks effect was highly significant, F (6, 216) = 16.39, $p < .001$. All three interactions with the blocks effect were significant beyond the .001 level. Examination of Fig. 1 suggested that these interactions were probably restricted to the initial three blocks of Phase 2. Therefore, two separate analyses of variance with repeated measures were performed on the mean run speed for the first three blocks and the last four blocks of Phase 2. The results of the first test showed that both the schedule and delay factors were significant, F (1, 36) = 17.86, $p < .001$, for schedule, and F (1, 36) = 4.46, $p < .05$, for delay, as was the Schedule × Delay interaction, F (1, 36) = 18.33, $p < .001$. The blocks effect was also significant, F (2, 72) = 3.94, $p < .05$. Both the schedule and delay factors interacted significantly with the blocks factor, F (2, 72) = 5.14, $p < .01$, and F (2, 72) = 4.64, $p < .05$, respectively. Tests of simple main effects showed that *S*s in the continuous delay group ran significantly slower than comparable *S*s in the continuous reward group, F (1, 36) = 20.44, the partial reward group, F (1, 36) = 20.08, as well as *S*s in the partial delay group, F (1, 36) = 36.18, all ps $< .001$. No other comparisons were significant.

Because of the interaction of blocks with both delay and schedule, F tests of simple effects were performed. It was found that the continuously rewarded *S*s showed a greater performance decrement than the partially rewarded *S*s, F (2, 72) = 9.95, $p < .001$, whereas no significant difference over blocks was detected between the two delayed groups. Further, the speed of the continuously rewarded *S*s decreased at a faster rate than that of the continuously delayed *S*s, F (2, 72) = 8.53, $p < .001$, but no such difference was obtained between the two partial groups. Moreover, the continuous reward group showed a greater performance decrement not only relative to the partial reward group, but also relative to the partial delay group, F (2, 72) = 9.34, $p < .001$.

The results of the analysis of variance test on the data of the last four blocks of Phase 2 showed a significant schedule effect, F (1, 36) = 48.77, $p < .001$, as well as a significant delay effect, F (1, 36) = 6.46, $p < .05$. The Schedule × Delay interaction was not significant ($F < 1$). The blocks effect was clearly not significant ($F < 1$), indicating that the groups had reached a stable level of performance over the last 16 trials of Phase 2. None of the ensuing interactions was significant, F (3, 108) = 1.13, for the Schedule × Blocks interaction, and $F < 1$, for both the Delay × Blocks and the Schedule × Delay × Blocks interactions. Individual comparisons

showed that the continuous reward group ran significantly slower than both the partial reward group, $F\ (1,\ 36) = 23.98$, and the partial delay group, $F\ (1,\ 36) = 45.48$, both ps $< .001$. No significant difference was found between the continuously rewarded and the continuously delayed *S*s, thus disconfirming the graphical "depression effect" which appears in Fig. 1. Similarly, the two partial groups did not differ significantly from one another, $F\ (1,\ 36) = 3.41$, $p > .05$. Finally, the continuously delayed *S*s ran significantly slower than either the partially rewarded *S*s, $F\ (1,\ 36) = 9.90$, $p < .005$, or the partially delayed *S*s, $F\ (1,\ 36) = 24.94$, $p < .001$.

The results of a simple two-way analysis of variance performed on the data of Block 7 showed that both main effects were significant, $F\ (1,\ 36) = 25.46$, $p < .001$, for schedule, and $F\ (1, 36) = 4.11$, $p < .05$, for delay. The interaction effect was not significant ($F < 1$). Individual comparisons among the groups revealed the same relationships obtained over Blocks 4–7. Both partial groups were significantly superior to the continuous reward group ($p < .001$), as well as to the continuous delay group ($p < .01$). No significant differences ($p > .05$) were obtained between the continuous delay and continuous reward groups or between the partial delay and partial reward groups.

A further inspection of Fig. 1 reveals that both partial groups as well as the continuous delay group showed very little change over Phase 2 relative to their terminal level in Phase 1. The performance of each group on each block of Phase 2 was compared with the group's mean performance over the last three blocks of Phase 1. By Block 2, the continuous reward group was already running significantly slower than its terminal level, $t(9) = 4.31$, $p < .01$, and continued to do so until the end of Phase 2. The continuous delay group showed no significant decrease over the entire phase. With the exception of the last block, both the partial reward and partial delay groups showed no significant decrease in performance. However, in Block 7, both groups ran slower than their terminal level in Phase 1, $t(9) = 3.10$, $p < .02$, for the partial delay group, $t(9) = 2.89$, $p < .02$, for the partial reward group.

Discussion

The results of the present study give strong support to the two hypotheses entertained earlier. First, positive transfer was clearly demonstrated when under conditions of continuous delay of reward, *S*s previously trained under partial reward or partial delay of reward conditions ran significantly faster than *S*s previously trained under continuous reward conditions. Moreover, this effect was rather stable as evidenced by the lack of any significant change in speed over the last 16 trials of the transfer phase. Second, a more or less complete transfer was obtained between nonreward and delay of reward as shown by the nonsignificant difference in running speed between the two partial groups in Phase 2.

To explain the PRE, Amsel (1958, 1962) proposed that frustration aroused by nonreward at first disrupts the ongoing goal-approach response, but later anticipatory frustration responses and their feedback stimuli (commonly referred to as the $r_f - s_f$ mechanism), instead of evoking avoidance responses as before, become counterconditioned to the ongoing response and actually enhance the instrumental response. In extinction, when both the continuously and partially reinforced *S*s experience frustration as a result of nonreward, the latter *S*s continue to run, or persist, longer in extinction than the former *S*s because the partially reinforced *S*s would have already been conditioned to make the instrumental response in the presence of frustration cues. In keeping with Amsel's frustration theory, experience with delay would be assumed to produce unconditioned avoidance responses (R_D) like the unconditioned frustration responses of nonreward (R_F). At first, anticipatory delay would evoke avoidance or interfering responses but, like the $r_f - s_f$ mechanism, a similar mechanism, $r_d - s_d$, would be formed and mediate the instrumental response. Under conditions of continuous delay, all *S*s would experience R_D. The *S*s in the continuous delay and partial delay groups would have already learned through the $r_d - s_d$ mechanism to make the instrumental response in the presence of delay cues. Therefore, both groups should eventually perform at about the same level.

A similar prediction can be made for *S*s in the partial reward and continuous delay groups, on the assumption that delay and nonreward are governed by similar underlying mechanisms. If nonreward could be seen as the limiting case of delay of reward (cf. Logan, 1960), then the previous assumption is quite tenable, and moreover, leads to the prediction of a complete transfer between frustrative nonreward and frustrative delay of reward, as was obtained in this study.

In conclusion, the present study has revealed three major findings. First, a sustained positive transfer was obtained between nonreward and delay of reward as shown by the continued superiority for 28 trials of both the partial reward and partial delay groups over the continuous reward groups. Second, a rather complete transfer was demonstrated since under conditions of continuous delay of reward, the group that previously received experience with nonreward on a random half of its trials was not significantly different from a group that previously received delay on a random half of its trials. Third, the positive transfer obtained was relatively permanent since both partial groups maintained their superiority over the base-line (continuous delay) group during the entire transfer phase. By the end of Phase 2, however, the two partial groups showed a tendency to converge toward the level of the base-line groups. It would seem that valuable information would be obtained if studies investigating transfer effects, as well as the PRE, incorporated in the design control *S*s which receive the test conditions from the beginning of training as was done in this study (cf. Daly, 1969). With such control *S*s, more meaningful statements can be made about the nature of transfer between persistence-producing operations.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin*, 1958, **55**, 102–119.

AMSEL, A. Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension. *Psychological Review*, 1962, **69**, 306–328.

AMSEL, A. Behavioral habituation, counterconditioning and persistence. In A. Black & W. K. Prokasy (Eds.), *Classical conditioning II*. New York: Appleton-Century-Crofts, in press.

BANKS, R. K. Persistence to continuous punishment and nonreward following training with intermittent punishment and nonreward. *Psychonomic Science*, 1966, **5**, 105–106.

BROWN, R. T., & WAGNER, A. R. Resistence to punishment and extinction following training with shock or nonreinforcement. *Journal of Experimental Psychology*, 1964, **68**, 503–507.

DALY, H. B. Learning of a hurdle-jump response to escape cues paired with reduced reward or frustrative nonreward. *Journal of Experimental Psychology*, 1969, **79**, 146–157.

DONIN, J. A., SURRIDGE, C. T., & AMSEL, A. Extinction following partial delay of reward with immediate continuous reward interpolated at 24-hour intertrial intervals. *Journal of Experimental Psychology*, 1967, **74**, 50–53.

GLAZER, H., & AMSEL, A. Transfer of persistence effects from frustration produced by blocking to frustration produced by nonreinforcement. *Psychonomic Science*, 1970, **18**, 311–312.

LOGAN, F. A. *Incentive*. New Haven: Yale University Press, 1960.

(Received April 29, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 103–109

EFFECTS OF CONTEXT ON LOWER ORDER RULE LEARNING IN SEQUENTIAL PREDICTION[1]

MARI RIESS JONES[2]

Ohio State University

Component-learning performance of five groups ($n = 16$) of college students was compared in contexts where sequential patterns were constant (C_1, C_2, C_3) or changing (R_1, R_2) arrangements of six run lengths. Order of occurrence of four event lights and nature of run arrangements were also manipulated. Run-component accuracy at common certainty points was a function primarily of pattern constancy and, to lesser extent, light order. The particular nature of a run arrangement had little effect on this component-learning measure, but did affect trials to a training criterion. Both training and transfer performance in a related four-choice task supported an interpretation that recurrent patterns facilitate component learning via reducing uncertainty about light positions and so decreasing memory-disruptive errors at light transition points.

How do people learn sequences of events? Do people attend initially to gross regularities or rhythms in a pattern as Simon and Kotovsky (1963) or Neisser (1966) suggest, or do they concentrate immediately on component properties such as lengths of runs (Myers, 1970; Restle, 1967)? The *S*'s initial approach to a sequential pattern is a topic that separates theorists who emphasize primitive perceptual processes from those who describe sequential learning as systematic rule acquisition. Reber (e.g., 1969), for example, noting parallels between sequential learning and language acquisition, suggests that *S* learn primarily syntactical relationships rather than strings of explicit symbols. But, Restle has maintained that with binary patterns, *S* first learns an inventory of run stems or lower order, mandatory rules. Thus, the pattern, AABAAA, which has runs of Event A of Lengths 2 and 3 is composed of a run stem of Length 2, which is learned initially with a mandatory rule of A → AA and two optional rules dependent upon this run stem (AA → AAA and AA → B). Others (e.g., Myers, 1970) also emphasize that *S* attends initially to the events within a run rather than to periodicities represented by arrangements of runs or of light positions. The role of higher order rules in mediating component learning is not clearly specified in either of these latter two descriptions. Jones (1971) has reviewed these issues in more detail.

Recently, Restle and Brown (1970) demonstrated the need for a theory involving higher order rules by showing that organized serial patterns were learned more rapidly than disjointed ones and that the use of lower order regularities (e.g., trills) did not depend upon certain higher order rules (Restle & Brown, 1970, p. 294). This theory has in common with Restle's (1967, p. 323) analysis the proposition that rule learning proceeds from lower order rules to higher order combinations.

The present study was designed to explore the relationship between higher order rule context and run-stem acquisition in order to define more clearly the manner in which *S* responds both to runs and to higher order relationships early in learning. Five sequences used in this study involved different arrangements of run lengths in two different light-order contexts. The manner in which higher order regularities could affect mandatory rule learning in such sequences is suggested by Myers (1970). Myers assumes that *Ss*

[1] This research was supported by a grant from the College of Social and Behavioral Sciences of Ohio State University. The author wishes to thank John O'Hara, Marvin Schreer, Andrew Sparks, and Steve Zartman for assistance in the collection and analysis of data.

[2] Requests for reprints should be sent to Mari Riess Jones, Department of Psychology, Arps Hall, Ohio State University, Columbus, Ohio 43210.

count the repetitive events in runs and that miscounting errors result from disruptive effects of prior errors at optional rule transition points. While Myers' theory does not attribute differences in transitional errors to higher order properties, presumably the mechanism of disruptive errors could reflect the effects of such contextual properties on component learning early in training.

Specifically, if *S* attends to sequential properties such as run arrangements, light order, or pattern constancy initially and if these do affect *S*'s ability to learn lower order rules, then mandatory rule learning should be poorer in contexts where these higher order properties are unavailable. In this study, early performance on mandatory rules was compared in contexts in which higher order properties were either manipulated systematically or were unreliable. If *S* attends to component rules initially, then changes in higher order rules should not affect early lower order rule performance. But if *S* does attend to and use higher order cues immediately, then these cues could affect component learning in one of two ways. Immediate sensitivity to contextual cues could furnish *S* with a specific overall structure that would afford *S* greater resistance to disruptive effects of short-term memory for component codes. This interpretation suggests that early in training, both accuracy and the conditional probability of an error on Trial n given one on Trial $n-1$ [$P(e_n|e_{n-1})$] would differ as a function of contextual manipulations. Alternatively, *S*'s sensitivity to higher order cues may simply permit him to narrow the set of alternative run codes appropriate to a particular event light. In this case, the effects of higher order rules on run predictions may become apparent only as *S* learns to use such rules to decrease uncertainty about the various light positions. Any degrading effects of context on component learning would result directly from the difficulty *S* has in learning higher order transitions. A further assumption that transition errors disrupt short-term memory for the current run permits the prediction that any effects of context revealed by run predictions will be accompanied by no differences in terms of $P(e_n|e_{n-1})$. Conditional error probability would decline only after accuracy at transition points increased sufficiently to permit accurate encoding of some runs into long-term memory.

In order to examine *S*'s use of context early in training, all *S*s were trained to a criterion of accuracy on run lengths only and were then transferred to a related four-choice sequence. This is consistent with the assumption of Myers (1970) that *S*s encode runs only at transition points. Higher order sequential properties were varied in three ways. First, the order in which *S* saw four event lights was manipulated so that the onsets of the lights, A, B, C, D (left to right), were either in Light Order (cycle) ABABCD or BABACD. Second, the arrangement of six run lengths ($k = 1, 2, 3, 4, 5, 6$, where k is the number of successive onsets of a given light) within a light cycle was either constant (C_1, C_2, and C_3) over repetitions of a cycle or changed haphazardly (R_1 and R_2). Thus, *S*s in C_1 and those in R_1 received the same run lengths for Light Order ABABCD, but for the former group the runs always occurred in the order 3–4–1–2–5–6, while, for the latter, run arrangements changed over repeated presentations of this light order. A restriction on the R_1 and R_2 conditions was that Light A occurred only in runs of 1 and 3 and Light B only in runs of 2 and 4. Condition C_3 represents the third way in which rule learning was examined. The pattern of run lengths in this condition was consistent with a highly learned sequence of integers. Presumably, if *S*s count events in runs and integer-code their lengths, this pattern should be easily extrapolated and learned. Table 1 outlines the five conditions.

If either light order, constancy of run pattern, or nature of run pattern affects component learning, then *S*s in these conditions should differ in their predictive accuracy at certainty points corresponding to mandatory rules which all five conditions have in common. The common certainty points of interest are those involving points at which a given run continues with pro-

TABLE 1

Sequence Descriptions, Total Performance Measures, and Component Learning Measures for the Five Training Conditions

Training cond.	Sequence descriptions: Light orders and associated run lengths	Total performance		Performance at common certainty points					
		Average total errors	Cycles to criterion	General error probability		Continuation-error probability		Perseverative-error probability	
				n	p	n	p	n	p
C_1	Lights A B A B C D Runs 3 4 1 2 5 6	41.5	8.31	511	.224	256	.143	183	.358
C_2	Lights B A B A C D Runs 2 1 4 3 5 6	50.5	9.94	589	.205	282	.133	270	.386
C_3	Lights A B A B C D Runs 1 2 3 4 5 6	60.8	12.75	730	.209	408	.150	243	.311
R_1	Lights A B A B C D Runs 1,3 2,4 1,3 2,4 5 6	55.6	8.31[a]	624	.313	333	.208	243	.416
R_2	Lights B A B A C D Runs 2,4 1,3 2,4 1,3 5 6	66.1	9.94[a]	727	.264	346	.173	300	.408

[a] Cycles to criterion for Cond. R_1 and R_2 were matched with those of C_1 and C_2, respectively.

ability equals 1.00 (run continuations), and those at which a run terminates with probability equals 1.00 (run terminations).

Method

Subjects.—The *Ss* were 96 female undergraduate students enrolled in introductory psychology at Ohio State University. Each *S* was randomly assigned to one of six experimental conditions ($n = 16$).

Apparatus.—The *S*'s booth contained a box with a front panel set with a row of four different-colored event lights, four corresponding response buttons, and a white center-mounted signal light. Although the four event lights were not labeled on *S*'s panel, they are referred to in this study as Lights A, B, C, D in order from left to right.

Control apparatus was located in an adjacent room. A Lafayette eight-bank program timer controlled the onset and durations of the event lights within a trial.

Design.—The five training sequences, each presented to a different group of *Ss*, are shown in Table 1. The training conditions differed with respect to: (*a*) two orders of occurrence of the four event lights (ABABCD and BABACD); (*b*) constancy of run arrangements over cycle repetitions (Cond. C_1, C_2, C_3 vs R_1, R_2); (*c*) arrangement of component runs within a constant run pattern (3–4–1–2–5–6; 2–1–4–3–5–6; 1–2–3–4–5–6). Arrangements of runs in Cond. C_1 and C_2 were mirror images of each other with respect to Lights A and B, while the run pattern of C_3 represented a highly learned number sequence. Sequences of R_1 and R_2 paralleled those of C_1 and C_2, respectively, in order of event-light occurrence, but the orders of run lengths in R_1 and R_2 were random with the restriction that Lights A and B occur with runs of 1 and 3 and 2 and 4, respectively, and Lights C and D with runs of Lengths 5 and 6.

The transfer sequence parallels the run arrangement of the training sequence of C_1 when those run lengths are multiplied by a factor of two. The order of light occurrence (DCDCBA) in the transfer task is a transformation of the order of lights in C_1. Training sequences of Cond. C_2, C_3, and C_1 are less directly related to the transfer task in terms of complex mediation via either light order (C_3, R_1) or run arrangement (C_2). The order of run lengths in the transfer pattern was 6–8–2–4–10–12; and these were associated with Lights DCDCBA, respectively. A sixth group of *Ss* received this transfer sequence with no pretraining (Cond. N).

Procedure.—The *Ss* were instructed to predict each light as rapidly and accurately as possible after the onset of the ready light on each trial. They were not informed of properties of the patterns, such as lengths of runs or run arrangements. The *Ss* had 1.2 sec. to respond after the onset of the .6-sec. ready light, after which there was an experimentally determined second interval before the signal for the next prediction. Each of the four event lights remained on for 1.1 sec. The *E* manually controlled trial onset and programmed the appropriate event light each trial.

In Cond. C_1, C_2, and C_3, *Ss* were trained until they had correctly predicted all run continuations within a pattern for two successive cycles. Accuracy of predicting run terminations, which are typically more difficult, was ignored, so that most *Ss* did not completely learn the training patterns. The *Ss* in Cond. R_1 and R_2 were equated with those in Cond. C_1 and C_2, respectively, in terms of number of cycle presentations. Because of the sequential order of runs in C_3, these *Ss* were treated separately.

Following training, the five experimental groups received the same transfer sequence. The former five groups were briefly apprised of a change in the

task. Transfer training was carried to a criterion of all correct predictions of the entire pattern for a single cycle. Those *S*s who could not meet this criterion were terminated after 15 transfer cycles (i.e., 630 trials).

Results

Training.—Table 1 presents different error probabilities at common certainty points across the five training conditions. For comparative purposes, all training errors were converted to probability scores. Continuation certainty points common to all conditions were after one and three B events, after two A events, and within runs of Lights C and D. The probability of an error in predicting these run continuations was recorded as a continuation-error probability for each *S*. Perseverative-error probability was based on repetition responses after three A's, four B's, five C's, and six D's. General error probability (P_G) at common certainty points included continuation errors, perseverative errors, and other run-termination prediction errors after longer run lengths. In all sequences, the correct event after the longer run lengths was predictable using knowledge of light orders within a cycle and the current run count.

Analysis of variance on P_G compared C_1, C_2 with random groups equated for number of training trials (R_1 and R_2). The main effect of run arrangement was significant, F (1, 60) = 13.48, $p < .001$, but the main effect of light orders was not, F (1, 60) = 3.00, $p > .05$. A more detailed analysis of errors revealed that differences between C and R groups were largely the result of continuation errors at common certainty points. Continuation error probability was significantly higher for R_1 and R_2 conditions than for C_1 and C_2, F (1, 60) = 7.91, $p < .01$. All groups were significantly less likely to make continuation errors on lights involving one run code (i.e., Lights C and D) than on those involving two (i.e., Lights A and B); the F for this main effect was F (1, 60) = 77.97, $p < .001$, and there was no significant Condition × Light Position interaction.

Perseverative-error probabilities over the four light positions did not differ significantly across the five conditions; again, all *S*s were less likely to perseverate erroneously on Lights C and D than after the long runs of Lights A and B, F (1, 75) = 50.51, $p < .001$. A Condition × Light Position interaction reflected the fact that *S*s in C groups were less likely to perseverate after the long A and B runs than *S*s in R conditions. A separate analysis of perseverative-error probability after the long A and B runs indicated *S*s in the two conditions of Light Orders BABACD (Cond. C_2 and R_2) were more likely to perseverate at these points than other *S*s.

Continuation, perseverative, and P_G analyses, based on marginal error probabilities, reveal poorer component learning with variable run patterns. The fact that more errors tend to occur in R sequences at true uncertainty points may account for such differences in that transitional errors could cause subsequent errors. Differences in marginal error probabilities might not be accompanied by differences in $P(e_n|e_{n-1})$. Conditional error probability was determined for each *S* for the first and second half of training. Average conditional error probabilities for each condition are plotted in Fig. 1. The overall effect of constant-run arrangements was significant in a planned comparison, F (1, 75) = 19.13, $p < .001$, although the reduction in runs of errors caused by pattern constancy was most marked in the second half of training. This is reflected in a significant Conditions × Trials interaction, F (4, 75) = 5.55, $p < .001$. The most important point to note is that the effects of run-pattern constancy were not significant in the first half of the session. Light order affected responding immediately only in the R groups, F (1, 30) = 8.94, $p < .01$. The *S*s in all conditions showed significant declines in $P(e_n|e_{n-1})$ over trials, F (1, 75) = 151.54, $p < .001$.

Other variables contributing to learning were examined in the performance of the three C conditions, which did not differ significantly in overall error probability nor in total errors (see Table 1), although they did in cycles to criterion, F (2, 45) = 9.06, $p < .001$. The latter difference

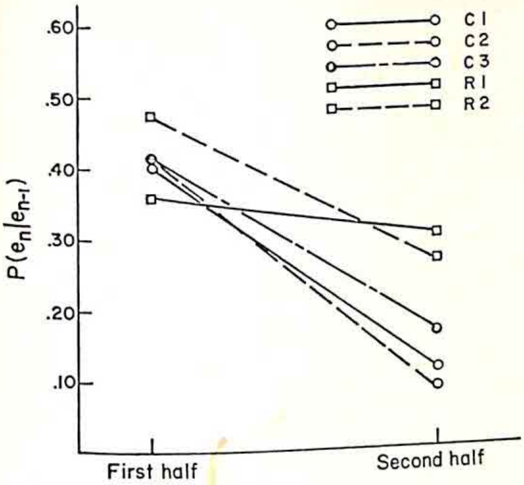

FIG. 1. The average probability of an error on Trial n given an error on Trial $n - 1$ over the first and second half of training for the five experimental conditions.

was due primarily to the greater difficulty of the sequence in C_3. A mean contrast of C_3 against the combined mean of C_1 and C_2 was unexpectedly significant in the wrong direction, $F(1, 45) = 7.84$, $p < .01$.

Differences among the three groups were small in terms of predicting Lights C and D and between continuation errors at those certainty points common to all five groups. Continuation-error and terminal error probability, P(e), after all six-run lengths for Cond. C_1, C_2, and C_3 are presented in Columns 5 and 6 of Table 2. Error probabilities for runs on Lights A and B are presented in order of ascending difficulty so that nonsignificant differences may be bracketed. Clearly, the probability of correctly predicting continuations of runs of A and B events after 1A or 2B, respectively, changes with context. Analyses of terminal error probabilities after runs of 1A, 3A, 2B, and 4B revealed no significant main effect due to sequence, but rather a significant Run Length × Sequence interaction, $F(6, 135) = 3.19$, $p < .005$. The run-length main effect was also significant, $F(3, 135) = 3.95$, $p < .01$. Results of contrast analyses on terminal error probability after runs on Lights A and B (Newman-Keuls) are outlined in Column 6 of Table 2 by vertical bars bracketing nonsignificant differences. It is clear from the pattern of results that the difficult points in the sequences are not necessarily those in which a run continues or branches in more than one way. In Cond. C_1, for example, after a single A event (Column 2, Table 2) two branches were possible: the next light could be either another A or a B. But after four successive B events, Light B could continue in only one way in Cond. C_1 by branching to Light A. This is because the arrangement of run codes for C_1 was 3A–4B–1A–2B–5C–6D. According to Restle's (1967) analysis, terminal error probability for 1A should be higher than that for 4B in C_1, and equivalent to that for 1A in other conditions. This is not the case. In terms of rank order of difficulty, run stems having two different branches (Column 3, Table 2) were easier than the other A and B runs in all conditions. Rather, the results suggest that run length, the relative position of the run in a pattern, and the maximum number of branches from a light position determine terminal error probability, although the manner in

TABLE 2

CONTINUATION AND TERMINAL ERROR PROBABILITIES FOR THE THREE CONSTANT RUN-ARRANGEMENT CONDITIONS

Cond.	Run stem/light	Branches/run stem	Maximum branches/light position	P(e) Contin-uation	Terminal[a]
C_1	1A	2	2	.1812	.4218
	2B[b]	2	3	.0752	.4789
	4B	1	3		.5612
	3A[c]	1	2		.5654
	5C	1	2		.4442
	6D	1	2		.3503
C_2	4B	1	2		.4540
	1A	2	3	.1491	.4739
	2B[c]	2	2	.2198	.5240
	3A[b]	1	3		.6403
	5C	1	2		.4111
	6D	1	2		.2569
C_3	3A	1	2		.3762
	1A[c]	2	2	.2464	.3842
	2B	2	3	.1058	.4112
	4B[b]	1	3		.5000
	5C	1	2		.3958
	6D	1	2		.3422

[a] Brackets in this column indicate nonsignificant differences.
[b] Fourth run in the cycle.
[c] First run in the cycle.

which these factors combine is not entirely clear.

Transfer.—The relative performance of all groups on the transfer sequence, which required *S*s to double the run lengths of the training sequence and systematically reverse light orders, demonstrates the superiority of *S*s trained with constant patterns. Only the main effect of run constancy was significant in the cycles-to-transfer criterion, F (1, 90) = 39.52, $p < .01$. No other comparison was significant, including that between the no-training control *S*s and *S*s who received the random sequences in training. In predicting both run onsets and continuations, *S*s pretrained with constant-run arrangements had fewer errors, F (1, 90) = 5.6412, $p < .05$, for continuation errors; F (1, 90) = 19.9942, $p < .01$, for run onset points, than *S*s in other conditions. Average total errors, average number of continuation and onset errors, and cycles to criterion are presented in Table 3. The *S*s in the three C conditions did not differ significantly in these measures, a finding which suggests again that the effectiveness of run constancy on code formation does not rest in *S*'s precise acquisition of the higher order rules themselves. All three groups were more accurate in predicting continuations of transfer light positions at which only one run length occurred (i.e., Lights A and B), despite the fact that these positions had been relatively more difficult in training, where either of two runs could occur, F (1, 40) = 30.57, $p < .001$. Thus, the current uncertainty of a particular light position is more effective in reducing average error than prior uncertainty of that light.

TABLE 3

AVERAGE NUMBER OF ERRORS OVER ALL OCCURRENCES OF LIGHTS A AND B, C AND D; AND AVERAGE NUMBER OF ONSET AND CONTINUATION ERRORS AND TRIALS TO CRITERION FOR ALL SIX GROUPS ON THE SAME TRANSFER SEQUENCE

Cond.	Transfer: Average total errors		Average errors		Cycles to criterion
	Lights AB	Lights CD	Continuation	Onset	
C_1	.659	1.465	.432	4.672	8.37
C_2	.505	1.006	.743	3.203	5.75
C_3	.560	1.175	.493	3.273	7.06
R_1	1.247	2.318	1.032	6.906	12.50
R_2	1.349	2.171	.966	6.531	11.18
N	1.494	2.709	1.064	8.234	12.94

Discussion

The comparatively low error probabilities within runs of like events at common certainty points is consistent with other data, establishing that *S* finds subunits of successively identical events highly salient (e.g., Myers, 1970). In other respects, the results are unequivocal in their support of the hypothesis that contextual cues affect *S*'s acquisition of mandatory rules that define run stems. Differences in error probabilities at common certainty points indicate that pattern constancy facilitates component learning. The effects of light order early in training were also apparent, primarily in preserverative-error probability in the constant-pattern conditions, suggesting that *S*s can effectively use this cue only when pattern constancy permits more accurate counting of events in the longer runs. Component learning was not affected by the particular run arrangement, despite the fact that the three different recurrent patterns did produce differences in trials to criterion.

No current theory of sequential learning can explain the entire pattern of results in the present study. The *S*s do attend to certain higher order regularities immediately, but the manner in which these properties affect component learning is not rule specific. The data are compatible with an interpretation that pattern constancy and light order operate on lower order learning primarily through reducing the likelihood of disruptive errors at transition points early in learning. This interpretation is supported by the finding that groups who differed in marginal accuracy at certainty points did not differ in conditional error probability initially. But while a transition error seemed to be equally disruptive in C and R conditions early in learning, it was less disruptive in the C conditions late in learning. As *S*s in the C conditions learned more about the pattern, they reduced not only transition errors, but general uncertainty about the run properties of the forthcoming event light. Thus, later in learning, these *S*s were less inclined to be disrupted by transition errors that

did occur. This analysis suggest that higher order and lower order rule learning of a pattern proceed simultaneously and interact.

Relationships between levels of rule learning are absent from both Restle's theories (1967, 1970) and Myers' (1970) run-distribution analysis. Neither predicts the differences in marginal and conditional error probabilities of the C and R conditions observed here.

Both approaches also fail to describe the relative learning performance of *S*s in the three constant-pattern conditions. The slower learning of *S*s presented with runs ordered as 1–2–3–4–5–6, for example, suggests that *S*s do not begin by attending to events in a run and coding the run length with an integer code. Rather, these *S*s seemed to be retarded in their learning as a result of interference by expectancies about alternations and left-to-right progressions. Garner and Gottwald (1967) have noted a similar phenomenon, and the finding is especially difficult for run-encoding theories. Also difficult for these theories was the complex interaction of run length, run position, and light-branching properties in the analyses of terminal error probabilities of these three conditions.

In contrast to these theories, perceptual learning theory tends to suggest that *S* attends initially to higher order relationships and to the nature of these relationships (e.g., Reber, 1969). The present study does not unequivocally support this interpretation. Particular run patterns did not specifically mediate run encoding, and neither run pattern nor light order resulted in positive transfer. Only *S*s pretrained with constant run patterns were more accurate in the test sequence. However, while it appears that pattern constancy is the primary contextual variable facilitating run encoding early in training, it can be argued that differences in the difficulty levels of the light-order variable were not sufficiently large in the present study to demonstrate mediation by this syntactic variable with a complex transfer sequence.

Within the constraints of the present design, the picture that emerges suggests that a simplified description of *S*'s hierarchical learning as beginning either with components (Myers, 1970; Restle, 1967) and working upwards or beginning with context (Reber, 1969) and then discriminating components is inadequate. Such descriptions ignore complex interactions of rule learning.

REFERENCES

GARNER, W. R., & GOTTWALD, R. L. Some perceptual factors in the learning of sequential patterns of binary events. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 582–589.

JONES, M. R. From probability learning to sequential processing: A critical review. *Psychological Bulletin* 1971, 76, 153–185.

MYERS, J. L. Sequential choice behavior. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*. Vol. 4. New York: Academic Press, 1970.

NEISSER, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1966.

REBER, A. S. Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, 1969, 81, 115–119.

RESTLE, F. Grammatical analysis of the prediction of binary events. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 17–25.

RESTLE, F., & BROWN, E. R. Serial pattern learning. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*. Vol. 4. New York: Academic Press, 1970.

SIMON, H. A., & KOTOVSKY, K. Human acquisition of concepts for sequential patterns. *Psychological Review*, 1963, 70, 534–546.

(Received May 3, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 110–114

EFFECTS OF BIMODAL STIMULUS PRESENTATION ON TRACKING PERFORMANCE [1]

DONALD H. McGEE AND RICHARD E. CHRIST [2]

Kansas State University *University of Portland*

The effects of unimodal and bimodal stimulus presentation were compared in a visual step function tracking task. The *S*s in Cond. VN were given a short burst of white noise simultaneous with the change in position of the visual input. In Cond. VT, *S*s received a short burst of one of four different tones when the target changed positions, each tone corresponding to a particular position in the visual display. In Cond. V, no auditory input was employed. The results indicate that a redundant auditory input can either facilitate or inhibit certain aspects of performance. These data were consistent with a perceptual model for bimodal stimulation.

Most previous research comparing unimodal stimulus presentation with bimodal presentation of correlated stimuli have shown evidence for intersensory facilitation; better performance is produced when stimuli are presented in two modalities as opposed to only one. This result has been found with reaction time experiments (Bernstein, Clark, & Edelstein, 1969a; Swink, 1966), vigilance experiments (Buckner & McGrath, 1963), and detection studies (Brown & Hopkins, 1967; Loveless, Brebner, & Hamilton, 1970). Handel and Buffardi (1969) have also shown that simultaneous auditory and visual stimulus presentation produces a faster rate of pattern identification than does individual modality presentation.

The present experiment was designed to explore the effects of bimodal, correlated stimulation on a complex motor task—tracking a repeating series of step function inputs. The distinguishing characteristic of this type of task, relative to those used in most previous bimodal studies, is that skilled tracking performance requires anticipation and the temporal-spatial organization of responses (e.g., Noble & Trumbo, 1967). The rationale for this investigation was that methods developed for analyzing tracking performance would allow specific statements concerning the locus of performance changes resulting from bimodal stimulation. That is, if bimodal stimulus presentation affects tracking performance, it would be possible through an analysis of oscillographic records to locate where the additional redundant input makes its contribution.

Method

Subjects.—Thirty-six right-handed, male college students volunteered to serve as *S*s for the experiment. These students were all between the ages of 18 and 25 and had normal vision and hearing. They were paid for their services.

Apparatus.—The apparatus was the Versatile Electronic Tracking Apparatus (VETA) developed at Kansas State University. Since a detailed technical description of the apparatus is available elsewhere (Trumbo, Eslinger, Noble, & Cross, 1963), only a general description will be presented here.

The VETA was used to present a repeating pattern of visual inputs on a pursuit tracking display. The input function, or target, was punched on mylar tape, read out by a Digitronics Model 2500 tape reader, converted to analog voltages by means of a digital-to-analog converter, and displayed on a cathode-ray tube (CRT) as a vertical line which moved along the horizontal axis of the CRT. The position of a similar vertical line, the cursor, was continuously adjustable by means of an arm control attached to *S*'s chair in a position for easy manipulation by *S*'s right arm. This arm control consisted of a horizontal arm rest, pivoted at the elbow on a vertical shaft, and an adjustable hand grip. A potentiometer attached to the lower end of the rotating shaft converted the arm control position into a continuously variable voltage which, in turn, drove the cursor on the CRT. A movement of 5.6° of the arm

[1] This article is based on a thesis submitted by the first author to Kansas State University in partial fulfillment of the requirements for the MS degree. The research was supported by the United States Air Force Office of Scientific Research, Office of Aerospace Research, under Contract F 44620-68-C-0020.

[2] Requests for reprints should be sent to Richard E. Christ, Department of Psychology, University of Portland, Portland, Oregon 97203.

control caused the cursor to move 1 cm. Both the target line and the control line were 16 mm. long and they overlapped by 2 mm.; the cursor appeared below the target.

Since both the input function and the control positions were represented as voltages within the system, the absolute difference between these two voltages was integrated over each trial by an operational amplifier manifold. This integrated error was read out on a digital voltmeter at the end of each trial. In addition, the input function and control positions were continuously recorded on an oscillograph during selected trials for purposes of analytical scoring.

An auditory input function was also punched on mylar tape and read out by a second Digitronics tape reader which was pulsed by the same timer used for advancing the visual input function. The auditory input consisted of either white noise generated by a Grason-Stadler Model 455C noise generator or four pure tones generated by four audio oscillators.

Two *S*s were run simultaneously in identical experimental booths. Each *S* was seated approximately 71 cm. from the CRT. The loudspeaker used to present the auditory input was about 50 cm. directly behind *S*'s head.

Inputs and conditions.—The visual input consisted of a step function in which the target moved from position to position in discrete jumps and remained at each position for .8 sec. The target moved through each of four different positions, then through the same four positions in a different order, defining an eight-element pattern. This pattern was repeated six times per trial so that each trial lasted 38.4 sec. The pattern in terms of voltages was −3, 7, 2, −6, 7, −3, −6, 2, where −7 would be the extreme left position and 7 the extreme right. Zero volts was at the center of the CRT. The equivalent of a 1-v. displacement on the display was 5.7 mm. Condition V consisted of *S*s who tracked this visual pattern without any simultaneous auditory input.

The different auditory inputs used in conjuction with the visual input produced two further conditions. In Cond. VN, the visual pattern of inputs was presented along with 40-msec. bursts of 50-db. white noise. The onset of the noise bursts occurred simultaneously with the changes in target position on the CRT. In Cond. VT, the visual pattern was accompanied by a synchronous pattern of tones. Four different tones were used to correspond to the four positions in the visual pattern. Each tone was 40 msec. in duration with loudness subjectively equated to that of the white noise. The frequencies of the tones were 300 Hz., 690 Hz., 1,590 Hz., and 3,650 Hz. The tones were arranged such that the 3,650-Hz. tone was presented with the extreme right position in the visual pattern (7), the 1,590-Hz. tone with the position 2, the 690-Hz. tone with the position −3, and the 300-Hz. tone with the extreme left position (−6).

Procedure.—The *S*s were randomly assigned, 12 each, to the three experimental conditions. For all conditions, *S*s were instructed to track the target on the CRT by keeping their cursor superimposed on the target as much as possible. They were told that there would be an eight-element pattern repeated six times in each of 30 trials and that the most accurate means of tracking this pattern would be to move simultaneously with the target. All *S*s were explicitly told that they should learn the pattern in order to perform well. Each trial was preceded by a 5-sec. warning light and followed by a 15-sec. rest interval. All 30 trials were presented in one experimental session. In the VT condition, *S*s were informed that four tonal bursts would be presented successively and that they would correspond to various target positions. The *S*s in Cond. VN were instructed that a brief burst of white noise would occur when the target changes position. All *S*s were informed of their integrated error score at the end of each trial.

Performance measures.—The global performance measure was absolute error integrated over individual trials. In addition, measures of temporal and spatial performance were obtained from the oscillographic records of the second repetition of the pattern on each of Trials 2, 4, 6, 10, 20 and 30. Temporal performance was determined by comparing the time that *S* initiated his primary movement relative to the target displacement; a movement initiated before target displacement is called a lead and a movement started after target displacement is called a lag. Measurements of starting time were made to the nearest 50 msec. Spatial performance was determined for primary movements which were associated with any lead or a lag of less than 150 msec. Primary movements associated with a lag greater than 150 msec. were not used for measuring spatial performance since the amplitude of these movements could be guided by the visual input rather than by *S*'s correct anticipation of the input function. Spatial performance is called an overshoot if *S* moved his control too far before stopping his primary movement and an undershoot if *S* did not move far enough before stopping. Spatial performmance was measured to the nearest 1 mm.

Results

Integrated error

Figure 1 shows integrated error as a function of the 10 three-trial blocks for each of the three conditions. These three-trial block data were used in an analysis of variance which showed tracking conditions, $F\ (2,\ 33) = 7.75$, and blocks, $F\ (9,\ 297) = 157.27$, to be the only significant effects, both $ps < .01$. Inspection of the data shown in Fig. 1 reveals that all groups improve over blocks but that Cond. VT is consistently superior to Cond. VN which, in turn, is consistently superior to Cond. V. However, Newman-Keuls tests on the group means show the VT and VN conditions not to differ from one

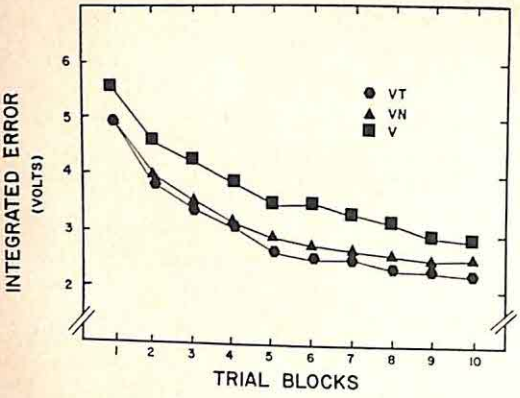

FIG. 1. Integrated absolute error as a function of trial blocks and input conditions.

another, but both differed significantly ($p < .05$) from Cond. V. It should be noted that the effects of auditory input (Conditions VT and VN) on integrated error are apparent as early as the first block of three trials. Integrated error data for only those trials used in analytical scoring were also examined, but these data produced results essentially identical to those shown in Fig. 1.

Analytic Scores

Temporal performance.—The measures of starting time were used to derive several different indexes of temporal performance. As an overall index of temporal error, the mean absolute starting time was determined for each *S* for each stage of practice. An investigation of these data showed that during the first 10 trials Cond. VN was more accurate (116-msec. starting time) than either Cond. VT (160 msec.) or Cond. V (152 msec.), but that by Trial 20 all three conditions were essentially identical (149 msec.). An analysis of variance of absolute starting times showed, however, that there was no significant main effect for conditions or trials. Likewise, there was no effect due to the Trial × Condition interaction.

A more revealing description of the effects of stimulus conditions on temporal performance was provided by a detailed look at the distribution of starting times. For the purpose of analysis, the overall distribution of starting times was divided into those that were beneficial in terms of demonstrating accurate anticipations of target displacements (beneficial anticipations) and those that demonstrated either an overanticipation (long leads) or a lag sufficiently long to suggest the lack of anticipation (long lags). Beneficial anticipations were defined as movements initiated within 150 msec. of the target displacement; long leads and long lags were defined as starting times greater than 150 msec. before and after target displacement, respectively. Table 1 shows the proportion of starting times falling into each of these categories for each condition at each stage of practice. (Results similar to the ones shown were obtained when 50 msec. and 100 msec. were used as criteria for categorizing starting times.)

A look at Table 1 shows that there was considerable variability in starting times for all groups early in the session and that variability decreased with practice due primarily to a decrease in the proportion of long lags. Furthermore, it can be seen that the proportion of beneficial anticipations was fairly constant over trials for all groups, while the proportion of long leads increased. Finally, if beneficial anticipations are used as an index of temporal accuracy, Cond. VN was more accurate over the first 10 trials (71% beneficial anticipations) than either Cond. VT or Cond. V (57% each).

Separate analyses of variance for each category of starting times tend to substantiate these visual interpretations of Table 1. The proportion of long lags were shown to decrease for all groups, $F(5, 165) = 30.50$, $p < .01$, but more so for Cond. V than for Cond. VT and VN, $F(10, 165) = 2.17$, $p < .05$. The proportion of long leads were affected only by trials, $F(5, 165) = 12.16$, $p < .01$. The analysis of variance of beneficial anticipations, however, found no significant effects. Analyses of the average magnitude of all leads and all lags, considered separately, and of the algebraic mean starting time yielded results which further substantiate these interpretations of temporal performance.

Spatial performance.—The measures of spatial performance were analyzed in the same manner as were the measures of

TABLE 1
DISTRIBUTIONS OF STARTING TIMES UNDER EACH CONDITION AS A FUNCTION OF PRACTICE

Trial	VT			VN			V		
	Long leads	BA	Long lags	Long leads	BA	Long lags	Long leads	BA	Long lags
2	21	50	29	14	65	21	5	44	51
4	23	59	18	8	79	13	12	64	24
6	28	65	7	19	72	9	16	71	13
10	40	55	5	26	67	7	30	53	17
20	34	66	0	38	61	1	39	56	5
30	45	55	0	38	62	0	34	64	2

Note.—Long leads and long lags refer to starting times greater than 150 msec. before and after target displacement, respectively. BA refers to beneficial anticipations; movements initiated within 150 msec. of target displacement. Data are in relative frequencies under each condition.

temporal performance. Figure 2 shows the average absolute magnitude of all spatial errors as a function of trials with input condition as a parameter. It can be seen in Fig. 2 that both bisensory conditions are associated with large spatial error on Trial 2 relative to Cond. V. It can also be seen that Cond. VN shows only a small, gradual improvement in spatial accuracy with practice, while Cond. VT shows a large, rapid improvement surpassing the performance of Cond. V which remains fairly constant. An analysis of variance of absolute spatial error showed significant ($p < .01$) effects for practice, F (5, 165) = 5.73, and for conditions, F (2, 33) = 5.45, but failed to show a significant Trial × Condition interaction. Newman-Keuls tests on the condition main effects showed that Cond. VT and V were essentially identical with respect to spatial error and that both were significantly superior ($p < .05$) to Cond. VN.

The overall distribution of relative spatial amplitudes was divided into three categories: large overshoots, spatially accurate movements, and large undershoots, where a primary movement terminating within 3 mm. of the target position was defined as a spatially accurate movement. The probability of a response falling into a given category of movement amplitude given a lead or a lag of less than 150 msec. was determined for each condition at each stage of practice. A visual inspection of these data showed that there was considerable variability in movement amplitude at each stage of practice. Conditions V and VT had a generally higher probability of a spatially accurate movement (.54) than Cond. VN (.39).

Separate analyses of the magnitudes of overshoots and undershoots, considered separately, and of the algebraic mean spatial error showed that all Ss tended to overshoot the target to a greater extent than they undershot. The magnitude of both types of spatial error decreased with practice, from 8 to 5 mm. for overshooting and from 7 to 4 mm. for undershooting. The relationships between conditions were essentially the same for the magnitudes of both overshoots and undershoots and are accurately reflected by the composite data shown in Fig. 2.

DISCUSSION

The results of this study indicate that redundant auditory input may or may not facilitate performance in a visual tracking task, depending upon which specific aspect of performance is measured and the type of information available in the auditory input.

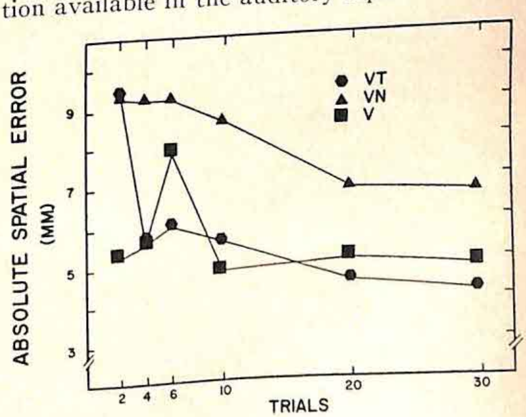

FIG. 2. Average absolute magnitude of spatial errors as a function of trials and input conditions.

In terms of global performance, measured by integrated absolute error, and in terms of the proportion of long lags, it was shown that the onset of an auditory input simultaneous with changes in visual input facilitated performance over that found without an auditory input. Furthermore, this facilitory effect occurs whether pure tones or white noise is used as the auditory input. These results, considered alone, would lend support to the energy integration model proposed by Bernstein, Clark, and Edelstein (1969b). This model assumes that stimulus intensities sum across modalities causing a joint auditory and visual event to be effectively stronger than a visual event alone. The energy integration model, however, was developed to account for data obtained in reaction time experiments. In the present study, the data suggest that *S*s are not generally reacting to the presentation of stimuli but are, instead, anticipating the input function. This anticipatory behavior is certainly evident by Trial 4 for the bisensory conditions, where less than 20% of the starting times could be classified as a typical reaction time.

Furthermore, a more thorough examination of performance in this tracking task raises several issues which cannot be handled with an energy integration explanation. If overall temporal accuracy, as measured by beneficial anticipations or by absolute temporal error, is considered, there is an initial tendency for the redundant noise condition (Cond. VN) to be superior to both the redundant tone condition (Cond. VT) and the vision-alone condition (Cond. V). These differences arise from the fact that *S*s in Cond. VT tend to initiate their responses too soon, while *S*s in Cond. V tend to initiate their responses too late. Furthermore, the additional noise input in Cond. VN was shown to interfere with spatial accuracy relative to Cond. V. and the additional tonal input in Cond. VT had no overall beneficial effect on spatial accuracy relative to Cond. V.

A more plausible explanation of these results would seem to require a perceptual or cognitive model, as opposed to the energy model proposed by Bernstein et al. (1969b) Such a model has been proposed by Handel and Buffardi (1969) to account for the data from their pattern identification study. In brief, they propose that different types of information may be contained in the logically similar inputs to different modalities. Hence, performance in a task using pairs of modalities may depend on how the information presented in the two modalities combine. For example, it is possible that the redundant timing information provided by the noise in Cond. VN aided timing but in the process distracted *S* from learning and responding accurately to the spatial cues in the visual display. On the other hand, the redundant positional information provided by the tones in Cond. VT may have caused *S*s to move too early in the direction of the next position; *S*s in Cond. V had to hold back the initiation of their primary movement until they had processed the position cues available from the visual display.

In order to assess the adequacy of an informational account of intersensory facilitation and inhibition, further research incorporating the properties of various types of inputs would be mandatory. In addition, this study found that global performance was predominately determined by the frequency and magnitude of response lags and very little by temporal or spatial accuracy as such. This fact supports the contention, made earlier by Nobel and Trumbo (1967), that a detailed analysis of response organization is necessary to uncover the information-processing and response strategies that may develop as a result of task conditions.

REFERENCES

BERNSTEIN, I. H., CLARK, M. H., & EDELSTEIN, B. A. Effects of an auditory signal on visual reaction time. *Journal of Experimental Psychology*, 1969, **80**, 567–569 (a).

BERNSTEIN, I. H., CLARK, M. H., & EDELSTEIN, B. A. Intermodal effects in choice reaction time. *Journal of Experimental Psychology*, 1969, **81**, 405–407 (b).

BROWN, A. E., & HOPKINS, H. K. Interaction of the auditory and visual sensory modalities. *Journal of the Acoustical Society of America*, 1967, **41**, 1–7.

BUCKNER, D. N., & MCGRATH, J. J. A comparison of performances on single and dual sensory mode vigilance tasks. In D. N. Buckner & J. J. McGrath (Eds.), *Vigilance: A symposium*. New York: McGraw-Hill, 1963.

HANDEL, S., & BUFFARDI, L. Using several modalities to perceive one temporal pattern. *Quarterly Journal of Experimental Psychology*, 1969, **21**, 256–266.

LOVELESS, N. E., BREBNER, J., & HAMILTON, P. Bisensory presentation of information. *Psychological Bulletin*, 1970, **73**, 161–199.

NOBLE, M., & TRUMBO, D. The organization of skilled response. *Organizational Behavior and Human Performance*, 1967, **2**, 1–25.

SWINK, J. R. Intersensory comparisons of reaction time using an electropulse tactile stimulus. *Human Factors*, 1966, **8**, 143–145.

TRUMBO, D., ESLINGER, R., NOBLE, M., & CROSS, K. A versatile electronic tracking apparatus (VETA). *Perceptual & Motor Skills*, 1963, **16**, 649–656.

(Received May 12, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 115–119

SPEED–ACCURACY TRADE–OFF WITH DIFFERENT TYPES OF STIMULI [1]

JAMES J. LYONS AND GEORGE E. BRIGGS [2]

Ohio State University

In a stimulus-classification task, linear function relating reaction time to central processing uncertainty were obtained under four conditions: high versus low accuracy, with letters or random figures as stimuli. The intercepts (not the slope constants) varied with accuracy level, while the slopes (not the intercepts) varied with stimulus material. This suggests that stimulus sampling rate is independent of stimulus familiarity but that central processing (stimulus classification) is slower for the less familiar stimuli.

Sternberg (1966) developed a methodology which permits one to determine additive components of reaction time and to do so without using the questionable insertion assumption required in the Donders (1868) methodology. The *S* memorizes a positive set of, say, one, two, or four stimuli and then makes one of two possible responses to each test stimulus: "yes," the test stimulus matches one of those items in the positive set, or "no," there is no match (the test stimulus is a member of the negative set). Typically, one finds either a linear relationship between reaction time (RT) and memory load (M): $\mathrm{RT} = a + b(\mathrm{M})$ or a linear relationship between RT and a Shannon expression of central processing uncertainty (H_c): $\mathrm{RT} = a + b(H_c)$, where H_c is determined primarily by memory load (see Swanson & Briggs, 1969). In either case, the additivity statement has been interpreted with reference to three sequential stages in human information processing: The slope constant b is interpreted as the time required per test to carry out the stimulus-classification functions at a central processing (second) stage; the intercept constant a represents the time required to perform the initial stage functions of stimulus encoding, sampling, and preprocessing plus the time to carry out a third stage which may be identified as response decoding.

Swanson and Briggs (1969) used random figures as stimuli in the Sternberg (1966) task and found that the equation $\mathrm{RT} = a + b(H_c)$ provided a satisfactory fit to the data and that the intercept constant a was linearly related to accuracy as the latter was expressed by the information transmitted (H_t) metric: $a = c + d(H_t)$. Thus, the more complete statement of additivity is

$$\mathrm{RT} = c + d(H_t) + b(H_c),$$

which when fitted to the data provided the following time constants:

$$\mathrm{RT} = .243 + .157(H_t) + .108(H_c).$$

In their discussion, Swanson and Briggs (1969) interpreted d as the time per bit of accuracy to carry out a stimulus sampling function such as Sperling (1967) suggested via his SCAN operator. The reciprocal of this constant (6.4 bits/sec) would be the rate of gain of information across SCAN under this interpretation. The reciprocal of the slope constant b (9.2 bits/sec) presumably is an estimate of the average rate of central processing functions involved in classification of the test stimulus into the positive or negative set.

Subsequently, Briggs and Swanson (1970) expanded upon the earlier study, and they used the more familiar letters of the alphabet as stimuli. Interestingly, they found an almost identical rate of stimulus sampling (6.7 bits/sec) to that in the

[1] This study was sponsored in part by the National Science Foundation through Grant GN–534.1 from the Office of Scientific Information Service to the Computer and Information Science Research Center, Ohio State University.

[2] Requests for reprints should be sent to George E. Briggs, Human Performance Center, Ohio State University, 404–B West 17th Avenue, Columbus, Ohio 43210.

earlier study; however, there was a faster rate estimated for the central stimulus-classification functions (13.5 bits/sec). This led Briggs and Swanson to the tentative conclusion that "while the rate of stimulus sampling may be independent of stimulus material, the speed of the subsequent central processing is a function of the familiarity and/or complexity of such material [p. 307]."

The present experiment was designed as a direct test of this hypothesis. According to the additive factor concept of Sternberg (1969), if the relative emphasis on speed affects the encoding stage of information processing and if the familiarity of the stimulus material affects the central processing stage, then a factorial comparison of two levels of accuracy and two levels of stimulus familiarity should reveal no interaction of these two variables. The impact of accuracy level should be on the intercept constant of the basic RT–H_c additivity statement; however, the constant a should not vary as a function of the stimulus material used. Moreover, the slope constant should vary with stimulus familiarity, but it should not be affected by changes in the accuracy of performance.

Method

The basic task was modeled after that of Sternberg (1966), described above. In all, three positive stimulus sets were used which differed in size as well as in the particular items included: $M = 1$, 2, or 4 items.

Combination of two types of stimulus material, letters versus random figures, with two levels of desired accuracy, high (95% correct responses) versus low (85% correct responses) defined four experimental conditions. In the letters condition, the seven positive set stimuli were chosen from the 14 letters of the alphabet used by Briggs and Swanson (1970, Exp. II). The remaining 7 letters were used as the negative probes. In the random figures condition, 14 of the eight-sided random figures used by Swanson and Briggs (1969) formed the stimulus pool. For both types of material, the assignment of specific items to positive or negative sets and to memory load within positive sets was counterbalanced across Ss.

Forty-four Ss responded to an ad in the university newspaper to serve individually for three daily sessions. Each S received $1.25 per session, plus up to $2.00 incentive pay. The latter was awarded in accordance with a bonus system to be described below. Eleven Ss were assigned upon order of arrival to each of the four experimental groups formed by combining the two types of stimulus material with the two levels of accuracy. In the first session, however, Ss in the high- and low-accuracy conditions were treated identically: The nature of the task was explained, S was encouraged to respond as quickly and as accurately as possible, and he then performed the classification task for a block of 48 trials under each of the three memory load conditions.

On the second day, a visual feedback device and a payoff matrix for the appropriate accuracy condition were introduced. The feedback display before S contained four counters: fast correct, slow correct, fast error, and slow error. After each response, E activated the appropriate counter causing the point value displayed on that counter to increase. For Ss in the high-accuracy conditions, the counters increased in steps of 10 points for a fast-correct response, -5 points for slow correct, -5 points for fast error, and -10 for slow error. In the low-accuracy conditions, the payoff matrix differed only in that the cost of fast errors was reduced to -1 point. The counters accumulated points over the 48 trials of a memory load (positive set) series and were then reset. Points were redeemed at the rate of 1¢ per 15 points, an exchange rate chosen such that the maximum possible bonus was approximately $1.00 per day. All Ss were told that the criterion times for distinguishing between fast and slow responses were based on their previous day's performance. In fact, however, where indicated by performance on the first eight trials, the time criterion was adjusted without S's knowledge in an effort to obtain the desired accuracy level. The payoff system and criterion adjustments were chosen on the basis of an earlier pilot study which indicated that they would yield the desired separation in the performance of the high- and low-accuracy conditions.

As in the first session, one block of 48 trials was presented for each positive set during the second session, and an identical procedure was followed for the final session. The order of presentation of the positive sets was counterbalanced across Ss on the second and third sessions, with the restriction that each positive set was used once each session. The E considered the first and second days as training sessions, and only the data from the final session were treated in the analyses reported below.

Two Shannon information metrics were employed: information transmitted H_t and central processing uncertainty H_c. The H_t measure of accuracy was calculated for each S in the current experiment from a 2×2 table containing the probabilities that a positive or negative response was made.

The index H_c has the status of an independent variable and was determined by the uncertainty associated with the various possible outcomes of central processing. Memory load is the primary determinant of the number of possible outcomes: $n - 1$ of the possible outcomes are a match of the test stimulus with one of the items in the positive set, while the remaining outcome is a no-match

situation. In this experiment, no match occurred half of the time in all three memory load conditions, and across memory loads each of the items within a positive set was presented equally often. Thus, the following values of H_c were obtained: for $M = 1$, $H_c = 1.0$ bit; for $M = 2$, $H_c = 1.5$ bits; and for $M = 4$, $H_c = 2.0$ bits. Briggs and Swanson (1970) present a complete accounting of the computation of H_c, as well as a comparison of the results of these computations to those obtained using several other possible methods of expressing the memory load variable.

The test stimuli were back-projected on a ground-glass screen 65 cm. before *S* by a Kodak RA 950 random access projector. Each test slide appeared for 2.5 sec. with a 3.5-sec. interstimulus interval. An electronic counter accurate to 1 msec. was activated by stimulus onset and was terminated by either of two switch closures: "yes," a match between test and memorized stimuli had occurred, or "no" match. The forefinger of each hand was used for responding, and, across *S*s, the left-right orientation of response buttons was counterbalanced. During each block of trials, *S* heard 80 db. of masking white noise, which was momentarily interrupted 1 sec. before the onset of each test slide. The *E* recorded the reaction time and accuracy for each response.

Results

Median reaction times were computed for each *S* across the last 40 trials in each memory load block of the third session. The H_t calculations were made for the same trials individually for each *S*, and frequency of correct responses was tabulated across *S*s for each experimental condition.

Considering first the high versus low accuracy data, it is evident that *E* was successful in obtaining a reasonable approximation to the accuracy levels desired: Overall averages of 96.3% versus 85.8% correct were obtained for the letters groups, and 95.3% versus 87.6% correct responses were recorded for the random figures groups. Chi-square analyses of these data indicate that the effect of accuracy level was significant ($p < .05$), while that of stimulus material was not. Finally, the effect of memory load on the accuracy data did not reach significance ($p > .50$), and so *S*s in each group can be considered to have worked at approximately equal accuracy across the three memory load levels.

The reaction time data were subjected to an analysis of variance. The three main effects (stimulus material, accuracy, and memory load) were significant at $p < .001$. In addition, of the four interaction terms, only the Stimulus Material × Memory Load interaction reached significance, $F(2, 80) = 3.71, p < .05$.

These results are displayed in Fig. 1, in which reaction time is plotted as a function of H_c. The parameter in Fig. 1 is information transmitted, averaged across *S*s and memory load. Reaction time is linearly related to H_c in all four cases. Further, since only the Stimulus Material

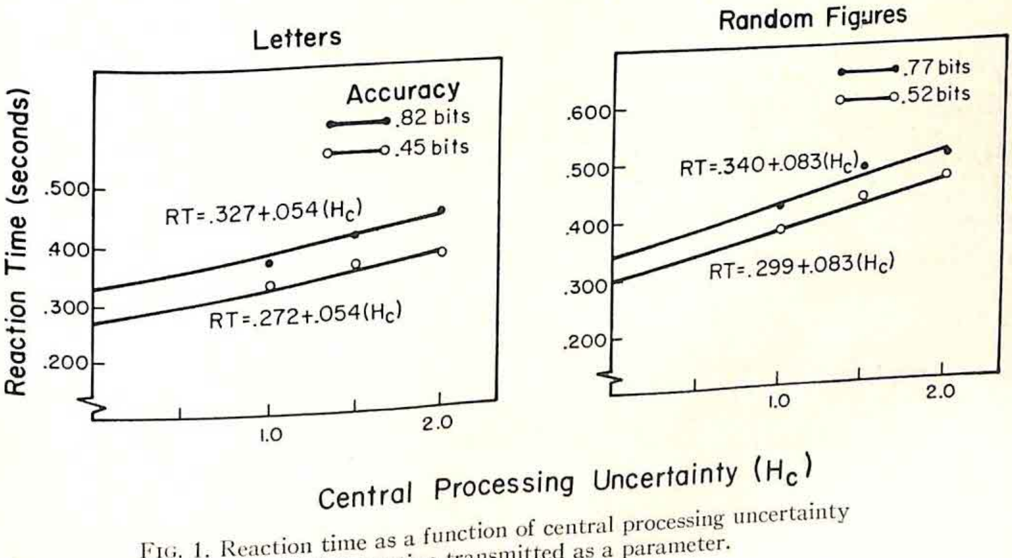

FIG. 1. Reaction time as a function of central processing uncertainty with information transmitted as a parameter.

× Memory Load interaction reached significance, parallel lines were fit separately to the data of Groups High-Letters and Low-Letters (see the left panel) and to the data of Groups High-Random Figures and Low-Random Figures (on the right). Best fits of the basic RT–H_c equation are shown.

First, note the effect of the accuracy variable within either of the two parts of Fig. 1. The only significant effect of increasing the accuracy of performance was to raise the intercept value of the equations. Next, compare the accuracy effect across the two parts of Fig. 1. The shift from low to high accuracy adds 55 msec. to the intercept for letters and adds slightly less (41 msec.) to the intercept for random figures. However, note the quantitative improvements in accuracy in going from the low to the high conditions in the two cases: It too is slightly more for letters (.37 bit) than for random figures (.25 bit). Thus, the magnitude of the increase in the intercept relative to the magnitude of the increase in accuracy was constant despite the differences in the type of stimulus material.

The type of stimulus material *did* affect the slope of the basic additivity equation: The slope of the equation fitted to the random figures data is steeper than that fitted to the letters data by 29 msec/bit. This is the significant Memory Load × Stimulus Material interaction reported above. Inspection of Fig. 1 also indicates that each of the intercept values for the random figures conditions is somewhat higher than the corresponding intercept for the letters conditions. The significance of this difference, however, is not directly available from the analysis of variance: The stimulus material main effect is based on the mean difference across memory load, not the difference at the point $H_c = 0$. A t test on the intercepts using pooled variance estimates was performed, and this indicated the mean difference of 20 msec. between the intercepts for letters and for random figures did not reach significance, $t(124) = 1.46$, $p > .05$, $SE = 13$ msec. The logic of this test is described in Draper and Smith (1966, Section 1.4).

Discussion

These results confirm Swanson and Briggs (1969) by demonstrating that trading speed for accuracy affects only the intercept of the function relating RT to H_c, not the slope constant. Following the Sternberg (1969) additive factor analysis, this result may be interpreted to mean that the processes involved in the speed-accuracy manipulations are independent of those whose durations vary with memory load. Presumably, both the stimulus encoding and the response decoding processes are of this nature and contribute to the magnitude of the intercept constant. Briggs and Swanson (1970) have shown, however, that the speed/accuracy effects are independent of response decoding effects; therefore, one may conclude that the locus of the speed/accuracy trade-off is in the initial encoding stage.

Given this, it is desirable to express the cost in encoding time of increasing the accuracy of performance. Following Swanson and Briggs (1969), the regression of the intercepts a from Fig. 1 on H_t was computed. Since, as noted, the Stimulus Material × Accuracy interaction was not significant at the intercepts, a single line was fitted to the four intercept values. The obtained equation expressing this decomposition of the intercept was $a = .211 + .154\ (H_t)$.

Now, 154 msec. per bit of information transmitted comes remarkably close to the previous estimates by Swanson and Briggs (1969) and by Briggs and Swanson (1970): 157 and 149 msec. per bit, respectively. Thus, the tentative conclusion reached by Briggs and Swanson from a cross comparison of several studies is confirmed within the context of the current, single experiment: The rate at which stimulus information is sampled appears to be independent of the type of stimulus material used. In accord with Swanson and Briggs, the following interpretation of the process of extracting information from the stimulus may be made: Immediately following presentation, a representation of the stimulus is held in a very short-term or "iconic" store. This representation contains more information than is needed for the task at hand. Therefore, S scans the icon, extracting information from it which is passed on for further processing. The current research, together with that of Swanson and Briggs and Briggs and Swanson, suggests that this scan operation gains information at a rate of approximately 6.5 bits/sec. Changes in relative stress on the speed of performance are

reflected in changes in duration, not the rate of this stimulus sampling operation.

So far we have focused on the finding that the encoding stage operations are not influenced by differences in the type of stimulus material used. It is not possible within the context of the present experiment to scale, or even to identify precisely all the key dimensions along which letters and random figures differ. However, it does seem reasonable to assume that many of the relevant dimensions are encompassed within the concept of "familiarity"; thus, one may consider the comparison of letters and random figures made here to be essentially a comparison of two levels of stimulus familiarity. It may be noted that Egeth and Blecker (1971) make similar use of the concept of familiarity. In their experiment, the actual types of stimulus materials viewed were (*a*) uppercase letters and (*b*) 180° inversions of those letters. They too found that matching responses were made more quickly to the more familiar material. However, their same–different task did not permit precise localization of the processing stage being affected by stimulus familiarity. The present experiment seems to provide the needed convergence, indicating that stimulus familiarity does not affect the rate at which stimulus information is sampled and preprocessed.

This is not to say, however, that stimulus familiarity is without effect. Return to Fig. 1 and consider the slope constants of the functions. Now, the slope constant reflects the change in reaction time as a function of the amount of *stored* information. In terms of an information-processing model, its reciprocal provides an estimate of the rate of operation of central processing. It is this stage which presumably includes memory search and comparison, as well as the associated retrieval operations, and it is here that we see a marked difference between the two types of stimulus material. Central processing of letters proceeded at an average rate of 18.5 bits/sec, whereas the same operations progressed at a substantially slower rate of 12 bits/sec when random figures were used. Thus, while the rate of sampling information from a stimulus appears independent of its familiarity, the subsequent speed of central processing apparently is directly influenced by the type of stimulus material used.

REFERENCES

BRIGGS, G. E., & SWANSON, J. M. Encoding, decoding, and central functions in human information processing. *Journal of Experimental Psychology*, 1970, **86**, 296–308.

DONDERS, F. C. Over de snelheid van psychische processen. Onderzoekingen gedann in het Psysiologisch Laboratorium der Utrechsche Hoogeschool. *Tweede Reeks*, 1868–1869, **2**, 92–120. (Trans. by W. G. Koster. *Acta Psychologica*, Amsterdam, 1969, **30**, 412–431.

DRAPER, N. R., & SMITH, H. *Applied regression analysis*. New York: Wiley, 1966.

EGETH, H., & BLECKER, D. Differential effects of familiarity on judgments of sameness and difference. *Perception & Psychophysics*, 1971, **9**, 321–326.

SPERLING, G. Successive approximations to a model of short-term memory. *Acta Psychologica*, Amsterdam, 1967, **27**, 285–292.

STERNBERG, S. High speed scanning in human memory. *Science*, 1966, **153**, 652–654.

STERNBERG, S. Discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, Amsterdam, 1969, **30**, 276–315.

SWANSON, J. M., & BRIGGS, G. E. Information processing as a function of speed versus accuracy. *Journal of Experimental Psychology*, 1969, **81**, 223–239.

(Received May 24, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 120–123

PROACTIVE INHIBITION IN FREE RECALL [1]

FERGUS I. M. CRAIK [2] AND JOHN BIRTWISTLE

Birkbeck College, University of London, England

Several workers have demonstrated that proactive inhibition (PI) reduces recall of verbal material in short-term memory (STM), and this finding has been taken as evidence for a continuum or one-process view of memory. In addition, the phenomenon of "release from PI" has led to conflicting views of the nature of encoding dimensions in STM. The present study shows that both PI and release from PI are confined to the secondary memory component of free recall and that primary memory is not vulnerable to these effects. This finding supports a two-process view of memory.

Keppel and Underwood (1962) showed that proactive inhibition (PI) operated in the short-term retention of verbal material. Using the Peterson distractor technique they demonstrated that there was virtually no forgetting on Trial 1, that forgetting increased over subsequent trials, and that this buildup of PI affected longer retention intervals (18 sec.) but not shorter intervals (3 sec.). Subsequent studies, summarized by Wickens (1970), have shown that "release from PI" is obtained if certain characteristics of the retained items are changed.

There are three points of theoretical importance which arise from this research. First, the phenomena provide a technique for exploring encoding dimensions in memory. Second, the finding of apparently identical PI effects in short-term memory (STM) and long-term memory (LTM) is a powerful argument against dichotomizing memory into separate stores. Third, the "buildup" and "release" phenomena may provide insights into causes of forgetting and into the manner of storage and retrieval of verbal material.

The present report focuses on the second issue. Melton (1963, 1970) has argued that the presence of PI in both short-term and long-term tasks is evidence for a one-process model of memory and against the STM/LTM dichotomy proposed by other theorists. However, an objection to this line of reasoning was put forward by Waugh and Norman (1965). They suggested that even immediate retention is mediated by two components—a short-term or primary memory (PM) component and a long-term or secondary memory (SM) component. Given that SM is involved in both short-term and long-term tasks, it it not surprising that some experimental manipulations affect performance in a similar fashion. Thus, before coming to a final conclusion on the dichotomy/continuity issue, it seems necessary to isolate the PM and SM components and see whether one or both processes are vulnerable to PI effects.

Tulving and Colotla (1970) provided a simple method for separating the PM and SM components in the free recall of word lists. They suggested that a retrieved word be classified as part of the PM component provided that no more than seven words (either later stimuli or responses) intervened between its presentation and its recall. Words with more than seven intervening items were considered to have been retrieved from SM. The purpose of the first experiment was thus to examine a typical free recall experiment for the presence of PI and to determine whether any such effects were located in PM or in SM.

Experiment I

Method.—Twenty student *S*s were presented with eight lists of 15 unrelated words for immediate free recall. Each list was read by *E* at a 2-sec. rate, and

[1] The authors would like to thank their colleagues in London for helpful discussion of this material. In particular, Exp. II was suggested by Elizabeth Warrington. Thanks are also due to Eva Gell for testing *S*s. The research was supported by a grant from the Medical Research Council.

[2] Requests for reprints should be sent to Fergus Craik, who is now at the Department of Psychology, University of Toronto, Toronto, Canada.

S then immediately wrote down as many words as he could remember in 1 min. Instructions were for free recall with the modification that *S*s were requested to always write down the last few words in the list first of all. The point of this modification was to equate *S*s' strategies and ensure that terminal words were retrieved from PM. The words used were drawn from a pool of 200 common two-syllable nouns. The *S*s were either tested individually or in small groups of 2 or 3. Each individual or group was presented with a unique set of eight 15-word lists drawn randomly from the common pool. After the task was explained, *S*s proceeded with the presentation and recall of the eight lists—no practice trials were given.

Results.—The total number of words recalled by each *S* from each list was broken down into PM and SM components. In this experiment a retrieved word was regarded as part of the PM component provided that no more than six items intervened between its presentation and recall. An analysis of variance on the data yielded a significant effect of trials, $F\ (7,\ 133) = 9.33$, $p < .001$, and a significant Trials $\times$ PM/SM interaction, $F\ (7, 133) = 11.25$, $p < .001$. Figure 1 shows that the significant interaction is due to the fact that PM recall remained constant at just over three words while the SM component declined fairly regularly from about five and one-half to two and one-half words. Thus the experiment has demonstrated the presence of PI in free recall and shown that this effect is located entirely within the SM component.

If PI in short-term recall is limited to SM, it seems likely that the release from PI phenomenon is also a property of SM. Experiment II investigated this proposition.

Experiment II

Method.—In an attempt to maximize PI, the words in successive free recall lists were all drawn from the same taxonomic class. The design consisted of giving one group of 20 *S*s five 15-word lists—all words being drawn from one class. A further group of 20 *S*s similarly received four homogeneous lists but on their fifth trial were presented with a list from a different class. The *S*s were tested individually. Each 15-word list was read by *E* at a 2-sec. rate and *S* then immediately wrote down as many words as he could remember in 45 sec. As in Exp. I, instructions were for free recall but *S*s were told to recall the terminal words first of all. Also, *S*s were given full details of the experiment beforehand—that is, one group was told that they would have five similar lists and the other group told, for example, "you will have four lists of animal names then one list of fruit and vegetable names." The category to be presented was given before each list. The point of these full instructions was to minimize any surprise or alerting effect of changing the category (Wickens, 1970). The words used were generated from common instances of the following classes: animal names, fruit and vegetables, parts of the body, natural phenomena, household objects, and trees and flowers. Each *S* received a unique set of lists drawn from the common pool.

Results.—The total number of words recalled from each list was divided into PM and SM components as in Exp. I. The mean numbers of words recalled under the various conditions are shown in Fig. 2.

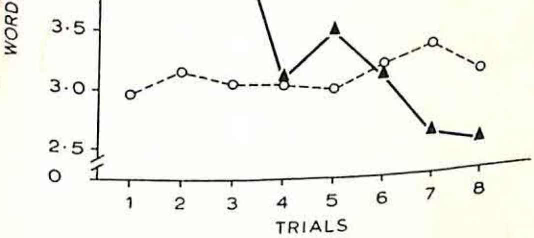

Fig. 1. The number of words recalled from PM and from SM as a function of trials.

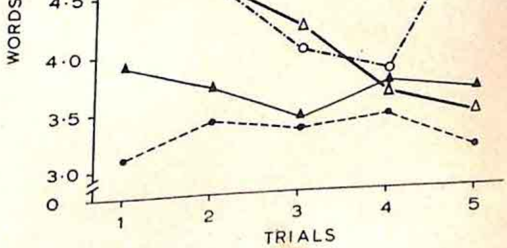

Fig. 2. The number of words recalled from PM and from SM as a function of trials and material.

This figure shows that the PM components remained relatively stable across the five trials. The SM components declined in a similar fashion for both groups of *S*s over the first four trials but while the "pure" PI group continued to show a decline on Trial 5, the "release" group showed a strong release from PI effect. These observations were supported by the results of analyses of variance performed separately on the PM and SM scores. The PM scores showed no effect of trials, ($F < 1$) while the SM scores showed a marked effect of trials, F (4, 152) = 9.65, $p < .001$, and a Trials × Groups interaction, F (4, 152) = 3.15, $p < .025$.

A second way of investigating the locus of the release effect on Trial 5 is to calculate the serial position curves for the PI group and the release group. The serial position curves, smoothed to reduce noise, are shown in Fig. 3. Positions 1 and 15 are the raw scores; other points represent the means of three successive positions. Figure 3 shows that the better recall by the release group is restricted to the first 10 serial positions; recall from the recency positions (generally accepted to reflect PM recall) is identical for the two groups. Again, it may be concluded that "release from PI" is an SM phenomenon. Figure 3 also makes the point that Exp. II generated typical free recall curves despite the slightly unorthodox instructions.

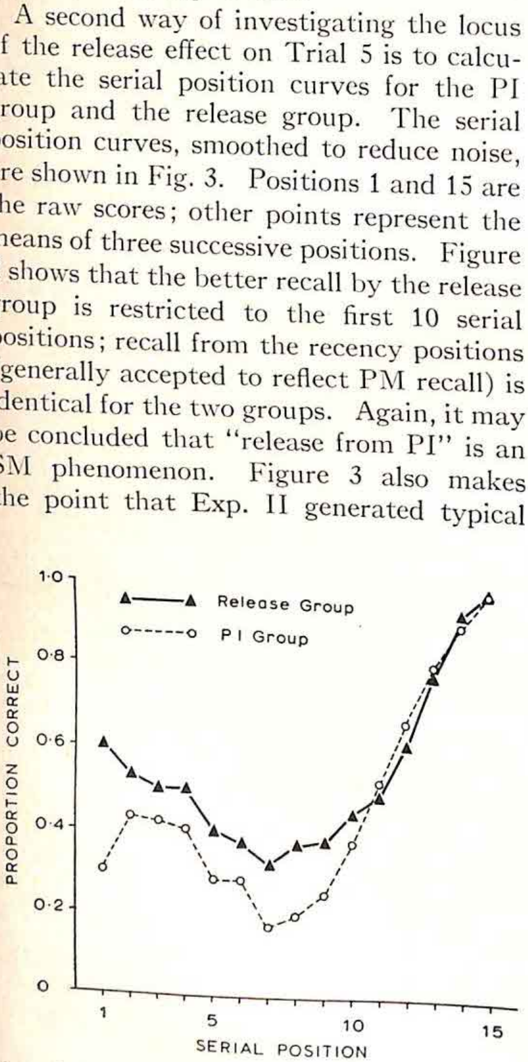

FIG. 3. Serial position curves for the "Release" and "PI" groups on Trial 5.

DISCUSSION

It has been shown that both the buildup of PI and the release from PI are effects limited to the SM component of free recall. The results directly confirm the suggestion of Loess and Waugh (1967) that items in PM are impervious to PI while items in SM are not. More generally, they are consistent with the view that "semantic" variables are confined to SM (Craik & Levy, 1970; Kintsch & Buschke, 1969).

The present demonstration used free recall, whereas most "PI in STM" studies have used the Peterson distractor technique. Can the present conclusions be extended to those earlier experiments? It seems likely that they can, since Glanzer, Gianutsos, and Dubin (1969) have shown that the PM component is eliminated by as few as six intervening items. Thus, since the PI and release from PI experiments typically use a 10 to 20-sec. retention interval, it may be argued that these studies are also showing effects in SM. Shorter retention intervals would presumably depend more on PM, and, by the present argument, should show smaller PI effects. In accordance with this prediction, both Keppel and Underwood (1962) and Loess (1964) found essentially no PI effects at a 3-sec. retention interval, but substantial effects at longer intervals.

Two points of theoretical interest follow from the present results. The first is that the presence of PI in short-term as well as in long-term retention is not good support for a continuum view of memory since a more detailed analysis of the short-term situation has revealed different effects of PI on the PM and SM components. In fact, the finding that terminal words in a free recall list are not vulnerable to PI effects while previous words are is a further argument in favor of two distinct processes operating in short-term recall. The second point is that if "PI" and "release from PI" are essentially dealing with the coding characteristics of SM, there is no need to attempt a resolution between the generally semantic type of coding in STM revealed by the release from PI studies (Wickens, 1970) and the generally nonsemantic view of PM suggested by other workers (Baddeley, 1966; Kintsch & Buschke, 1969). These two sets of studies are simply dealing with two different components of short-term retention.

The point of view that PI effects are restricted to SM does not necessarily imply that the effects are permanent or even long term in their duration. In fact, studies by Loess and Waugh (1967) and by Kincaid and Wickens (1970) have shown that PI dissipates within 2 min. using the Peterson technique. However, the point of view adopted here *does* imply that in "short-term PI" experiments, retrieval is from the associative-semantic component of memory.

Finally, do the present results say anything about the nature of PI or the release phenomenon? The most reasonable view at the moment seems to be that the buildup of PI is synonymous with a drop in the discriminability of stored items (or in the effectiveness of retrieval cues) as the experiment proceeds. When the category of stimuli is changed, the new dimensions of coding are thus less noisy and performance is enhanced. If retrieval from PM relies on the use of very transient temporal-acoustic cues (Tulving, 1968), then the finding that such cues are not rendered less effective by information from previous trials makes good sense.

REFERENCES

BADDELEY, A. D. Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, 1966, **18**, 362–365.

CRAIK, F. I. M., & LEVY, B. A. Semantic and acoustic information in primary memory. *Journal of Experimental Psychology*, 1970, **86**, 77–82.

GLANZER, M., GIANUTSOS, R., & DUBIN, S. The removal of items from short-term storage. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 435–447.

KEPPEL, G., & UNDERWOOD, B. J. Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 153–161.

KINCAID, J. P., & WICKENS, D. D. Temporal gradient of release from proactive inhibition. *Journal of Experimental Psychology*, 1970, **86**, 313–316.

KINTSCH, W., & BUSCHKE, H. Homophones and synonyms in short-term memory. *Journal of Experimental Psychology*, 1969, **80**, 403–407.

LOESS, H. Proactive inhibition in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 362–368.

LOESS, H., & WAUGH, N. C. Short-term memory and intertrial interval. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 455–460.

MELTON, A. W. Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 1–21.

MELTON, A. W. Short- and long-term postperceptual memory: Dichotomy or continuum? In K. H. Pribram & D. E. Broadbent (Eds.), *Biology of memory*. New York: Academic Press, 1970.

TULVING, E. Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

TULVING, E., & COLOTLA, V. A. Free recall of trilingual lists. *Cognitive Psychology*, 1970, **1**, 86–98.

WAUGH, N. C., & NORMAN, D. A. Primary memory. *Psychological Review*, 1965, **72**, 89–104.

WICKENS, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, **77**, 1–15.

(Received June 5, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 124–128

A SMALL-TRIALS PREE WITH ADULT HUMANS: RESISTANCE TO EXTINCTION AS A FUNCTION OF NUMBER OF N-R TRANSITIONS [1]

LEONARD POON AND JOSEPH HALPERN [2]

University of Denver

Human *S*s received 10 acquisition trials under conditions of partial reinforcement in a modified gambling task. Groups differed as a function of percentage of reinforcement and number of N-R transitions. In addition to the demonstration of a small-trials PREE, the results were in agreement with predictions made from sequential theory. Alternative theories developed from frustration models are also discussed.

A partial reinforcement extinction effect with small numbers of acquisition trials (SMPREE) has been successfully demonstrated in many studies with animals (Robbins, 1971). Although there is no clear agreement as to what defines limited acquisition training, McCain (1966) reports the SMPREE with as few as two to six trials. Only very limited attention has been focused on the SMPREE with adult humans, although such an effect presumably exists since Lewis and Duncan (1956) report a PREE with as few as eight acquisition trials. They modified a slot machine so that trial outcomes could be controlled. The *S*s were instructed to continue playing until they desired to stop. The dependent measure was number of trials to quitting. In addition to a clear PREE, Lewis and Duncan obtained the expected inverse relationship between number of trials to extinction and percentage of reward.

Recent sequential theories of instrumental conditioning (Capaldi, 1967) have minimized the effects of percentage of reward on resistance to extinction. According to Capaldi's (1967) sequential theory of instrumental conditioning, the PREE is determined by the distribution of Nonreinforced (N) and Reinforced (R) trials. In an experiment using rats and 10 acquisition trials, Spivey (1967) supported sequential theory. In that experiment, Spivey used four groups which differed in percentage of reinforcement and number of N-R transitions (number of times an R trial was followed by an N trial). In addition to a strong SMPREE, Spivey found that resistance to extinction was independent of percent of reinforcement but increased with increases in number of N-R transitions. Finally, Robbins (in press) in a review of the alleyway literature has concluded that Capaldi's sequential theory best handles the majority of the data relating to animal PREEs.

Halpern and Poon (1971) report a series of experiments designed to evaluate Capaldi's (1967) sequential hypothesis with adult humans. While alternative theoretical formulations were proposed, the results were generally consistent with sequential theory. In these experiments, acquisition training involved a minimum of 32 trials, and the research was limited to evaluating those portions of sequential theory which deal with specific N-length constructs. The present experiment employs a methodology similar to the one used by Lewis and Duncan (1956). The *S*s received 10 acquisition trials with percentage of reinforcement and number of N-R transitions varying between groups.

Method

Subjects.—The *S*s were 105 University of Denver undergraduate volunteers randomly distributed into seven groups. A total of 118 *S*s were run in order to obtain 15 *S*s in each of the seven groups. Thirteen *S*s were disqualified because of equipment failure, *E*'s errors, or because *S*s chose to stop playing during the acquisition phase.

[1] This research was supported by a faculty grant from the research fund of the Graduate School, University of Denver to the second author.

[2] Requests for reprints should be sent to Joseph Halpern, Department of Psychology, University of Denver, Denver, Colorado 80210.

Apparatus.—The *S* was seated on an executive swivel chair in front of a component integrated console. The console was arranged in a semicircle consisting of the following components: a gray box housing electrical relays labeled "random sequence programmer," an intercom, a modified Gerbrands nickel dispenser with a "coin acceptance" box and a "payoff" box, a raised panel labeled "electronic simulated slot machine" with a pilot light labeled "ready signal" affixed on the top center of the panel. On the bottom of the panel were two contact switches registering the decision of "play" or "no play." Another feedback light labeled "coin deposit indicator" was located in line with the "play" contact switch and feedback light. The *E* was situated in an adjoining room separated by a one-way mirror. A tape recorder was used for instruction along with an intercom to handle questions. Hunter timers were used for automatic timing of the gambling sequences. Control and display panels were situated in front of *E* for monitoring *S*'s choice behavior and for administering reinforcement.

Procedure.—The *S* was told that he was in a decision-making experiment. The electronic simulated slot machine was controlled by the random sequence programmer and *S* had a choice of "play" or "no play" on every trial. "No play" was defined as turning off the machine for a trial, and the choice of "no play" would result in an automatic penality of one penny. The *S* also had a choice of stopping play completely by simply pushing a "stop" button. It would cost *S* one nickel to play, and if he won the machine would pay off three nickels. He was told that he would keep his winnings at the end of the game.

The *S*s were randomly assigned to one of seven groups. Each group received 10 acquisition trials according to the schedules in Table 1. The *S*s in Group CR received continuous reinforcement while the remaining partial reinforcement (PR) groups are distinguished on the basis of percent of reinforcement (first unit of the group designation) and number of N-R transitions (second unit of the group designation).

Several criteria were employed as indicants of extinction. Pilot research has shown that *S*s typically cease playing after seven consecutive "no plays." Another criterion of extinction within the limits of the experiment was the number of trials before *S* ran out of money. The third criterion, which was used by Lewis and Duncan (1956), was the number of trials before *S* decided to stop playing. Resistance to extinction was simply measured by the number of trials before *S*s reached one of the above criteria of extinction.

TABLE 1

ACQUISITION SCHEDULES

Group	Trial									
	1	2	3	4	5	6	7	8	9	10
CR	R	R	R	R	R	R	R	R	R	R
30–1	R	R	N	N	N	N	N	N	N	R
30–3	N	N	R	N	N	R	N	N	N	R
50–1	R	R	R	R	N	N	N	N	N	R
50–3	N	N	R	R	N	R	R	N	N	R
70–1	R	R	R	N	N	N	R	R	R	R
70–3	N	R	R	R	N	R	R	N	R	R

RESULTS AND DISCUSSION

The mean number of trials to extinction for the CR group was 26.86, and the means for the PR groups are given in Table 2. A comparison between the mean number of trials to extinction for the CR and PR groups showed a considerable SMPREE, t (103) = 4.11, $p < .01$. More important are the differences between the PR groups. Inspection of Table 2 shows small differences among groups as a function of percent of reinforcement and more marked differences with respect to number of N-R transitions. A two-factor analysis of variance, with independent variables defined as percent of reinforcement (30, 50, and 70) and number of N-R transitions (1 and 3) with number of trials to extinction as the dependent variable, yielded only a significant main effect of number of N-R transitions, F (1, 84) = 25.419, $p < .01$. These results confirm Capaldi's (1967) sequential theory and replicate the findings of Spivey (1967).

Much of the strength of Capaldi's (1967) theory lies in the fact that it provides predictions which are counter to the typical findings. That is, the inverse relationship between percent of reinforcement and resistance to extinction, until recently, has been a ubiquitous empirical phenomenon and has formed the basis for

TABLE 2

MEAN NUMBER OF TRIALS TO EXTINCTION FOR ALL PR GROUPS

No. of N-R transitions	Percent of reinforcement		
	30	50	70
1	45.50	38.13	41.66
3	63.66	54.93	57.20

many of the so-called expectancy theories of the PREE. Consequently, more important than the gross effects shown by the analysis of variance are the simple effects obtained via comparisons of cell means. Six specific predictions can be formulated where the empirically determined inverse relationship between percent of reinforcement and resistance to extinction would yield predictions different from those provided by Capaldi's theory. The first set of three comparisons involves groups in which percent of reinforcement is constant and number of N-R transitions vary (30–3 vs. 30–1, 50–3 vs. 50–1, 70–3 vs. 70–1). Table 2 shows that in all three cases, resistance to extinction was greater for the group with three N-R transitions. The second set of three comparisons is obtained by pairing groups in such a way that the larger percentage of reinforcement appears with the larger number of N-R transitions (50–3 vs. 30–1, 70–3 vs. 30–1, 70–3 vs. 50–1). Again, from Table 2 it can be seen that the groups experiencing higher percentages of reinforcement are also most resistant to extinction. The chance probability of six specific directional predictions for six binary events is .015. An analysis of the six comparisons was performed with Duncan's new multiple-range test (Edwards, 1968). Significant differences were found for 30–3 vs. 30–1 ($p < .01$), 50–3 vs. 50–1 ($p < .05$), 70–3 vs. 70–1 ($p < .05$), and 70–3 vs. 70–1 ($p < .01$). Differences between the remaining two groups resulted in $p < .10$. The results, then, provide clear support for the sequential theory predictions of the SMPREE.

While the present results clearly support a sequential theory explanation of SMPREE, it is not equally clear that the responsible variable is the number of N-R transitions. In this, as in other limited acquisition experiments designed to test sequential theory (e.g., Spivey, 1967), there is some confounding between the number of N-R transitions and certain N-length variables, where N length is defined as the number of consecutive N trials preceding an R trial. The groups with one N-R transition experience a single long N length, while those with three N-R transitions have three shorter N lengths. The confounding is inevitable since it is impossible to devise a schedule in which N lengths are constant while N-R transitions and percent of reinforcement vary without simultaneously varying the number of acquisition trials. If one were to take Capaldi (1967) literally, with short acquisition (a situation where presumably no single N length occurs with a frequency sufficient to reach asymptotic habit strength), N-length frequency is more important than N-length magnitude. In one sense, then, the present results confirm the prediction that with preasymptotic habit strength N-length frequency rather than N-length magnitude is the critical variable. Unfortunately, it thus becomes impossible to determine whether N-length frequency or number of N-R transitions account for the present results.

Alternatives to Capaldi's (1967) sequential theory typically involve conditioning assumptions such as those incorporated in Amsel's (1958) frustration model. The model, as described by Amsel, Hug, and Surridge (1968), specifies that the SMPREE is dependent on experiencing at least one R trial prior to the N trials —nonreward is frustrating only in the presence of the fractional goal response (r_g) and r_g depends upon prior reward. Inspection of Table 1 shows that all groups with three N-R transitions initiated training with N trials while all groups experiencing one N-R transition began acquisition with R trials. Since each of the groups with three N-R transitions were more resistant to extinction than any of the groups receiving one N-R transition, our results are clearly contradictory to expectations from the Amsel, Hug, and Surridge (1968) frustration model.

A modified version of frustration theory for the SMPREE has been suggested by Brooks (1969), who emphasized the fact that unconditioned frustration (R_F) was greater following CR than PR. Since limited acquisition training experiments typically yield CR groups who experience more reward than PR groups, the former

groups enter extinction with more R_F. While Brooks has impressive support for the effect of R_F, Capaldi and Waters (1970) have handled Brooks' findings within sequential theory. Traupmann and Wong (1971) supported a counterconditioning hypothesis based on frustration theory by interpolating CR training between a PR acquisition phase and extinction. Two groups of rats were given either RRRR or NRNR and immediately extinguished, while in two other groups the above schedules were followed by 16 R trials prior to extinction. A PREE was noted only in the groups with interpolated extinction. The absence of a PREE in the groups without interpolated CR training can be attributed to any one of several factors. Specifically, the SMPREE tends to be a rather elusive phenomenon. Variables such as magnitude of reward (Surridge, Rashotte, & Amsel, 1967) and intertrial interval (McCain, 1968) can easily determine whether or not the effect will be present at all, especially with only four acquisition trials. Finally, it is not at all clear that when 16 CR trials are interpolated between PR and extinction, resulting in 20 acquisition trials, that one is dealing with an SMPREE.

While it may not be possible to speak directly to certain of the components of the theories developed by Brooks (1969) and Traupmann and Wong (1971), the possible effects of interpolated CRF training and equating for the number of rewarded acquisition trials are important for sequential theory. Thus, an additional group of 15 *S*s was run in exactly the same manner as Group 50–3 except that five R trials were interpolated between the schedule shown in Table 1 and extinction. This group thus experienced interpolated CRF training (Traupmann & Wong, 1971) and the same number of rewarded acquisition trials as the CR group. The mean number of trials to extinction was 57.96. This result suggests, at least for the present task, that the effect of equating the number of rewarded acquisition trials is virtually nil, as is the effect of interpolated training.

One final point deserves some discussion. Specifically, the definition of just what constitutes "small trials" is somewhat arbitrary. One criterion frequently used for a specification of small trials derives from frustration theory, i.e., fewer rewarded trials than necessary to produce frustration on N trials. Since one of the purposes of the present experiment involved an attempt at replication of Spivey (1967), 10 acquisition trials were used. It is, however, clear that our human *S*s have more information at the first acquisition trial (informed of win and lose contingencies) than do rats. Consequently, if one admits the viability of the frustration-derived specification of small trials, it is unlikely that we are dealing with a small-trials effect, although acquisition training was clearly limited. It may therefore not be appropriate to limit the generality of the present results to just small-trials phenomena.

REFERENCES

Amsel, A. The role of frustrative nonreward in noncontinuous situations. *Psychological Bulletin*, 1958, **55**, 103–117.

Amsel, A., Hug, J. J., & Surridge, C. T. Number of food pellets, goal approaches, and partial reinforcement effect after minimal acquisition. *Journal of Experimental Psychology*, 1968, **77**, 530–534.

Brooks, C. I. Frustration to nonreward following limited reward experience. *Journal of Experimental Psychology*, 1969, **81**, 503–505.

Capaldi, E. J. A sequential hypothesis of instrumental learning. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*. Vol. 1. New York: Academic Press, 1967.

Capaldi, E. J., & Waters, R. W. Conditioning and nonconditioning interpretations of small-trial phenomena. *Journal of Experimental Psychology*, 1970, **84**, 518–522.

Edwards, A. L. *Experimental design in psychological research*. (3rd ed.), New York: Holt Rinehart & Winston, 1968.

Halpern, J., & Poon, L. Human partial reinforcement extinction effects: An information-processing development from Capaldi's sequential theory. *Journal of Experimental Psychology Monograph*, 1971, **89**, 207–227.

Lewis, D. J., & Duncan, C. P. Effect of different percentages of money reward on extinction of a lever-pulling response. *Journal of Experimental Psychology*, 1956, **52**, 23–27.

McCain, G. Partial reinforcement effects following a small number of acquisition trials. *Psychonomic Science Monograph Supplements*, 1966, **1**, 251–270.

McCain, G. The partial reinforcement effect after minimal acquisition: Single pellet reward. *Psychonomic Science*, 1968, **13**, 151–152.

Robbins, D. Partial reinforcement: A selective review of the alley literature since 1960. *Psychological Bulletin*, 1971, **76**, in press.

Spivey, J. E. Resistance to extinction as a function of number of N-R transitions and percentage of reinforcement. *Journal of Experimental Psychology*, 1967, **75**, 43–48.

Surridge, C. T., Rashotte, M. E., & Amsel, A. Resistance to extinction of a running response after a small number of partially rewarded trials. *Psychonomic Science*, 1967, **7**, 31–32.

Traupmann, K. L., & Wong, P. T. P. The small-trial PREE and interpolated CRF training. *Psychonomic Science*, 1971, **23**, 207–208.

(Received June 5, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 129–135

HEMIRETINAL EFFECTS IN TACHISTOSCOPIC LETTER RECOGNITION [1]

D. O. NEILL, H. SAMPSON, [2] AND J. A. GRIBBEN

University of Auckland

The *S*s reported letters from unilateral and bilateral displays under all 15 possible hemiretinal viewing conditions administered in random order. Right visual field superiority was found from unilateral displays and left visual field superiority from bilateral displays. The data also showed that viewing with two hemiretinae is better than viewing with one hemiretina and that viewing with temporal hemiretinae is better than viewing with nasal hemiretinae. Comparisons of monocular, triple hemiretinal, and binocular viewing consistently demonstrated these results, and analysis of data collected under partial report instructions again showed superiority of temporal hemiretinal viewing.

Tachistoscopic studies investigating recognition of letter displays may involve stimulation of one or both eyes from one or both sides of a central fixation point in the visual field. When viewed binocularly, stimuli presented bilaterally (to both left and right sides of fixation) are more accurately recognized in the left visual field (LVF) than in the right visual field (RVF). However, for binocular viewing of stimuli presented only unilaterally, either left or right of fixation, superior recognition of stimulation in the RVF has typically been reported. Reading habits, attentional processes, and various types of dominance have been used to explain these results (see White, 1969). At least part of the reason for this difference in field superiorities when viewing unilateral and bilateral displays may be the order of reporting the stimulus display, since single-field (LVF or RVF) random report of *bilateral* displays (Fitzgerald & Marshall, 1967) and forced RVF-then-LVF report (Douglas, 1968; Freeburne & Goldman, 1969) yields RVF superiority.

Results from monocular viewing of letter displays sometimes closely resemble the well-established binocular results (Goodglass & Barton, 1963) or show similar field differences between each eye's visual fields with overall accuracy of one eye better than the other (Barton, Goodglass, & Shai, 1965; Koetting, 1970). But at least three studies have reported greater field differences for left-eye viewing of unilateral displays (Markowitz & Weitzman, 1969; McKeever & Huling, 1970; Overton & Wiener, 1966), with stimulation of the left eye's nasal hemiretina (left eye's LVF) leading to poorest accuracy of the four individual hemiretinae. This result may be accounted for by the joint action of left-cerebral and right-eye (acuity) dominance, or by temporal hemiretinal superiority with either cerebral or eye dominance.

Not many monocular studies of bilateral displays have been reported. Crovitz and Schiffman (1965) found no eye differences with their bilateral displays. Both Corballis (1964, two vertical letter pairs) and Crovitz and Daves (1962, in Crovitz & Lipscomb, 1963, 7° span of digits) found an RVF superiority for the right eye; and Crovitz and Daves also found RVF superiority with binocular viewing, suggesting that at least in the Crovitz and Daves study, the spacing of the stimuli, wider than the critical angle of about 4° (Kimura, 1959), has led to this atypical result.

The intention of the present study is to examine, within the same experimental sessions, all 15 possible viewing conditions: not only the 9 conditions of monocular and binocular viewing of unilateral and bilateral displays, but also the double nasal and double temporal hemiretinal viewing conditions, and the 4 triple

[1] Based on research by the senior author in partial fulfillment of requirements for the MS degree.

[2] Requests for reprints should be sent to H. Sampson, Department of Psychology, University of Auckland, Private Bag, Auckland, New Zealand.

hemiretinal viewing conditions that may provide information about additive effects in monocular–binocular comparisons. For all these 15 viewing conditions the same type of stimulus was used, and *S*s were asked to report the display in the same order, left to right. The conditions differed only in the number (0, 1, or 2) and type of hemiretina (nasal or temporal) stimulated by each visual field. On any trial the particular condition was always unknown to *S*. Comparisons of the various conditions should allow close examination of any changes in recognition that occur as stimulating conditions change and could allow general statements on viewing conditions that are more broadly based than those from monocular and binocular studies alone.

Method

Subjects.—Five psychology department staff or students served as *S*s. All were optometrically examined and accepted as *S*s with the following criteria: (*a*) full binocular vision with stereopsis, (*b*) normal ACA ratio and no marked heterophoria at the viewing distance of 1.22 m., (*c*) visual acuity equal to or better than 6/6 and equal acuity in each eye under the binocular-balance, simultaneous condition, (*d*) refractive status within the range from emmetropia to .75D hyperopia (equivalent sphere), equal in each eye and with no more than .50D astigmatism, (*e*) interpupillary distance of 60 ± 2mm., and (*f*) right-handed (dextrality index of at least .85, Humphrey, 1951).

Apparatus.—Two fields were used of an electronically programmed tachistoscope manufactured in the department. Each field was illuminated by white-coated, cold cathode lamps, reflecting .8-mL. luminance. The adaptation field contained a central black fixation spot of 4′ diameter. All stimulus exposures were 50 msec. The *S*s triggered the display themselves by pressing a microswitch. Specific combinations of hemiretinae were stimulated by altering combinations of vertically sliding occluders in a metal plate positioned .62 m. from *S* in the optical path of the stimulus field. The occluders had slanted edges to prevent light passing between them and allowed precise hemiretinal stimulation from each eye's visual field with an interpupillary distance of 60 ± 2 mm. The adaptation field with the fixation spot was seen binocularly, subtending 2°50′ vertically and 3°26′ horizontally, and viewing of the stimulus field (also 2°50′ × 3°26′ when viewed fully binocularly) was determined by the combination of occluders let down into the stimulus field optical path.

Stimuli.—Stimuli were six random, black, upper-case letters (Letraset 287), excluding M and W, spaced in two groups of three letters each side of the center of an off-white card. Individual letters subtended 14′ vertically and a mean 10′ horizontally at the 1.22-m. viewing distance. Each group of three letters subtended 45′ and the total six-letter display 2°15′. Each of the letters used occurred approximately equally often at each of the six positions. There were 75 stimulus cards, randomly sorted into five groups of 15 cards each.

Procedure.—The *S*s were instructed to maintain fixation on the central spot in the adaptation field, to trigger the stimulus when ready, and to verbally report the stimuli from left to right. The 15 viewing conditions were administered in five different random orders in one experimental session. Four similar sessions were run. On any trial, *S* could not determine what type of hemiretinal stimulation would occur or whether the display would be unilateral or bilateral. Individual letters were scored correct regardless of their location in the report sequence.

Results

Table 1 shows the mean percent correct recognition for the 15 viewing conditions, the results for each visual field shown separately. Diagrams on the left of Table 1 indicate which hemiretinae are stimulated in each condition, and the hemiretinae are named in the cells next to the percent correct recognition figure. *Binocular viewing* of these displays exhibited the usual findings, i.e., RVF superiority of unilateral displays (a difference of 9.0%, see Fig. 1, heavy line; Cond. 13 and 14) and LVF superiority of bilateral displays (a difference of 20.0%, see Fig. 2, heavy line; Cond. 15). Analysis of results from these conditions yielded a significant Visual Field × Unilateral/Bilateral interaction, $F\,(1, 4) = 12.38$, $p < .05$. *Monocular viewing* of *unilateral* displays led to RVF superiority only for left-eye viewing (left eye 22.0% RVF superiority, right eye 7.0% LVF superiority, see Fig. 1, thin lines; Cond. 1, 2, 3, and 4), Eye × VF interaction, $F\,(1, 4) = 12.18$, $p < .05$. This result resembles more closely that of Overton and Wiener (1966) than that of Markowitz and Weitzman (1969) or McKeever and Huling (1970), although these studies also show a greater laterality difference for the left eye. With monocular viewing of *bilateral* displays (Cond. 7 and 8), both eyes showed LVF superiority (see Fig. 2, thin lines) like the normal binocular

TABLE 1

DIAGRAMS OF CONDITIONS OF STIMULATION AND MEAN PERCENTAGE CORRECT RECOGNITION UNDER EACH CONDITION

No.	Diagram of stimulation from above	LVF	RVF
1	○ ⦸	83.7 RT	
2	○ ⊘		76.7 RN
3	⊘ ○		LT 83.0
4	⦸ ○	LN 61.0	
5	⊘ ⦸	69.0 RT	LT 57.3
6	⦸ ⊘	LN 63.0	53.0 RN
7	○ ⊗	77.0 RT	43.0 RN
8	⊗ ○	LN 59.0	LT 54.3
9	⊗ ⊘	LN 59.3	LT 64.7 RN
10	⊗ ⦸	LN 80.0 RT	LT 50.7
11	⦸ ⊗	LN 80.3 RT	47.0 RN
12	⊘ ⊗	71.0 RT	LT 61.3 RN
13	⦸ ⦸	LN 80.3 RT	
14	⊘ ⊘		LT 89.3 RN
15	⊗ ⊗	LN 76.3 RT	LT 56.3 RN

Note.—Blank cells indicate field occlusion. Named hemiretinae (LN, RT, LT, RN) in cells identify type of stimulation for that condition.

result, but the larger laterality difference was in the right eye (34.0% compared to the left-eye laterality difference of 4.7%;

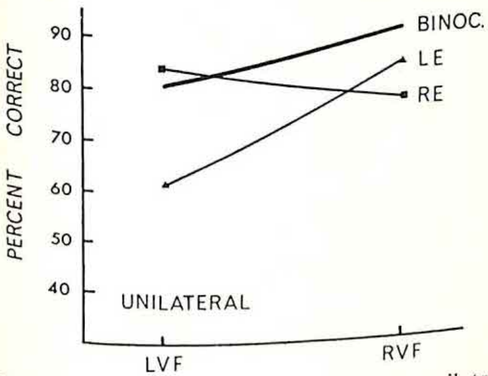

FIG. 1. Percent correct recognition of unilateral displays with binocular viewing (thick line) and monocular (LE and RE, thin lines) viewing.

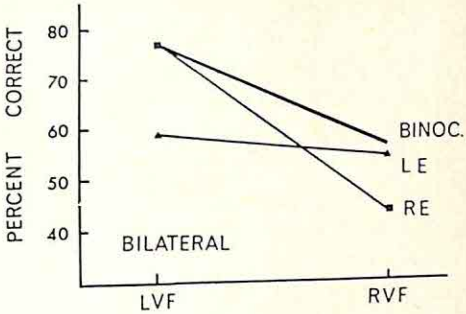

FIG. 2. Percent correct recognition of bilateral displays with binocular viewing (thick line) and monocular (LE and RE, thin lines) viewing.

Eye × VF interaction, F (1, 4) = 52.82, $p < .01$). Both these monocular unilateral and bilateral Eye × VF interactions are consistent with a significant overall temporal (compared to nasal) hemiretinal superiority and agree with the Markowitz and Weitzman finding of an overall temporal hemiretinal acuity superiority in a noncompetitive viewing situation.

Monocular–binocular comparisons.— Binocular viewing generally allowed more accurate recognition than monocular viewing (see Fig. 1 and 2). The binocular level of recognition, however, was close to the monocular level when monocular viewing involved temporal hemiretinal stimulation. This suggests that the added nasal hemiretinal stimulation in binocular viewing has only a small effect. Scheffé tests on contrasts of nasal, temporal, and both hemiretinal conditions in Fig. 1 and 2 showed that only the nasal hemiretinal conditions differed from the others at the .05 level of significance.

Double nasal and double temporal hemiretinal stimulation.—The effects of nasal and temporal stimulation may be relatively stable and be maintained in various combinations of stimulating conditions. A direct test of this stability can be made by comparing monocular viewing (Cond. 7 and 8) with double nasal and double temporal hemiretinal stimulation (Cond. 5 and 6), a comparison involving only single hemiretinal stimulation in each visual field, and the two groups of conditions differing only in the type of hemiretina that is being simultaneously stimulated. The left eye's

hemiretinae are fairly similar in this comparison (LN scores in monocular viewing, Cond. 8, vs. LN scores in similar hemiretinal viewing, Cond. 6, differing by 4.0%; and LT scores in monocular viewing, Cond. 8, vs. LT scores in similar hemiretinal viewing, Cond. 5, differing by 3.0%, see Fig. 3). The right eye's scores differ by 10.0% (RN) and 8.0% (RT). Stimulating the other eye's similar hemiretina simultaneously appears to lessen laterality differences, but the LVF superiority remains (10.8% overall, F (1, 4) = 15.50, $p < .05$) and does not significantly differ from the monocular viewing LVF superiority of 19.3% (Monocular/Similar Hemiretina × VF interaction, F (1, 4) = 2.38, $p > .10$). The smaller laterality differences may be the result of a minor compensatory effect: with similar hemiretinae stimulated at the same time, each being approximately equally efficient, the total amount of recognition is more evenly divided between each field.

Triple hemiretinal viewing.—Fig. 4 presents data together from monocular viewing and data from the triple hemiretinal conditions. The heavy line in Fig. 4a shows recognition percentages with left-eye viewing of bilateral displays, and in Fig. 4b with right-eye viewing of bilateral displays (the same data as plotted in Fig. 2). The added points in Fig. 4 are triple hemiretinal conditions which are monocular viewing conditions with one hemiretina of the other eye also viewing the bilateral display. Considering the left eye, the "extra" hemiretinal conditions are Cond. 9 and 10 (RN also viewing RVF, and RT also viewing LVF); for the right eye these "extra" conditions are Cond. 11 and 12 (LN also viewing LVF, and LT also viewing RVF). Points representing data from these conditions are identified in Fig. 4 by, for example, +RT, which indicates that in this case the LVF is seen by both the left eye and by the right eye's temporal hemiretina. Points in the other visual field, at the other end of the line representing triple hemiretinal conditions, represent data when that visual field is seen by only one hemiretina (the same hemiretina as the monocular viewing condition). Analysis showed that recognition is improved when a visual field is seen by two instead of one hemiretina while the other, simultaneously viewed visual field is seen by only one hemiretina, F (1, 4) = 17.66, $p < .05$, and is greater for the added temporal hemiretina, F (1, 4) = 34.08, $p < .01$, as a main effect in difference scores. Another analysis of triple hemiretinal conditions can be made by comparing them with binocular viewing conditions. Now, instead of observing an *additive* effect of triple hemiretinal viewing on monocular viewing, we may observe a *subtractive* effect of occluding one hemiretina's view of displays seen binocularly. In Fig. 5 the triple hemiretinal conditions are plotted so that the effect of occluding one hemiretina can be studied. The heavy line represents data of binocular viewing, and the added points are identified, in an analogous way to Fig. 3 by the hemiretina that is occluded. In Fig. 5a, the line with the LVF point labeled − RT indicates that it is the right eye's temporal hemiretinal that is occluded and that the data point at the other end of this line is from two-hemiretinal viewing. Recognition is impaired by occluding one hemiretina, F (1, 4) = 8.05, $p < .05$, and the impairment is greater for temporal hemiretinal occlusion, F (1, 4) = 8.40, $p < .05$.

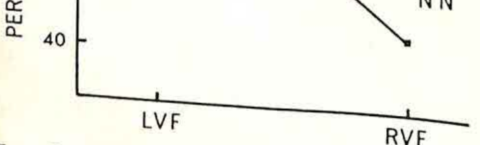

FIG. 3. Percent correct recognition of bilateral displays with monocular viewing (RE and LE, continuous lines) and similar hemiretinal viewing (TT and NN, broken lines, Cond. 6 and 5 of Table 1).

The comparisons of triple hemiretinal viewing (whether adding or occluding one hemiretina) with monocular and binocular viewing both lead to the same conclusions.

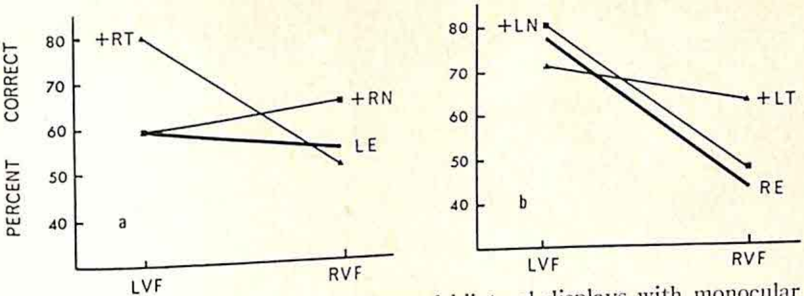

FIG. 4. Percent correct recognition of bilateral displays with monocular and triple hemiretinal viewing. (In Fig. 4a, two triple hemiretinal conditions are compared with left-eye viewing, identified by +RT (RT and LN viewing LVF and LT viewing RVF, Cond. 10) and +RN (RN and LT viewing RVF and LN viewing LVF, Cond. 9). In Fig. 4b, two triple hemiretinal conditions are compared with right-eye viewing, identified by +LN (LN and RT viewing LVF and RN viewing RVF, Cond. 11) and +LT (LT and RN viewing RVF and RT viewing LVF, Cond. 12).)

Recognition is better when viewing a display with two rather than one hemiretina, and better when viewing with a temporal rather than a nasal hemiretina. Both these conclusions are extensions of the comparison of monocular and binocular viewing of unilateral as well as bilateral displays. Triple hemiretinal stimulation does not unequivocally confirm any compensatory effect that may be present in the monocular and double nasal and double temporal viewing conditions. Although laterality differences obtained with triple hemiretinal viewing do seem to swing away from those obtained with binocular viewing (Fig. 5), a significant difference is found only when one hemiretina is occluded. Recognition in the other visual field does not show much compensatory improvement, $F(1, 4) = 4.64$, $.10 > p > .05$, and the correlation of *S*s' triple hemiretinal-binocular difference scores is only $r = -.18$ (mean within-cell correlation). Comparison of triple hemiretinal viewing with monocular viewing (Fig. 4) is even less convincing. Here, recognition in the visual field seen by only one hemiretina remains fairly stable and quite similar to the monocular result.

Compensatory effects in "other visual field" conditions can be examined by comparing *all* single hemiretinal viewing results when the other visual field is either not seen or seen with either one or two hemiretinae. When this analysis is made, apart from a difference associated with unilateral and bilateral displays, there was nothing found that has not been described earlier (i.e.,

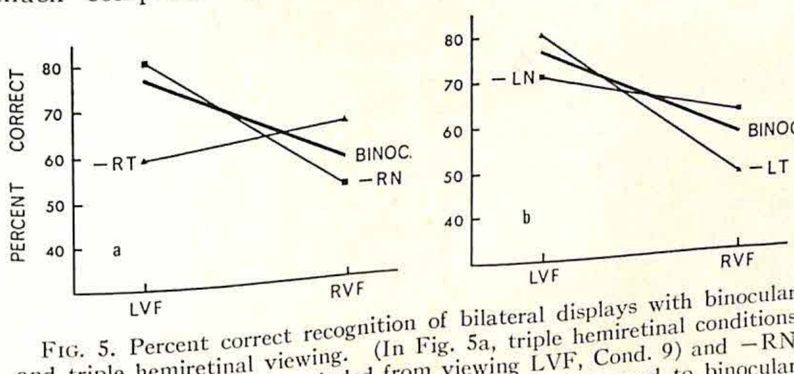

FIG. 5. Percent correct recognition of bilateral displays with binocular and triple hemiretinal viewing. (In Fig. 5a, triple hemiretinal conditions indicated by −RT (RT occluded from viewing LVF, Cond. 9) and −RN (RN occluded from viewing RVF, Cond. 10) are compared to binocular viewing. In Fig. 5b, triple hemiretinal conditions indicated by −LN (LN occluded from viewing LVF, Cond. 12) and −LT (LT occluded from viewing RVF, Cond. 11) are compared with binocular viewing.)

temporal greater than nasal, smaller laterality difference for similar hemiretinal viewing, VF × Unilateral/Bilateral display interaction). A difference in recognition between unilateral and bilateral single hemiretinal conditions may, of course, be regarded as a compensatory effect. Unilateral scores are higher than scores from one visual field in bilateral displays. However, this difference is associated with reporting three versus six letters, and so percent correct scores will obviously be higher when only three letters are to be reported. Much the same conclusion applies when all two-hemiretinal viewing conditions are compared.

Partial report.—Fifteen additional trials were run under the single hemiretinal conditions (Cond. 1–12) when the five *S*s were required to report only one visual field, that field indicated by a thick bar above and below one side of the displays. Under partial report instructions, RVF superiority was universal. In the bilateral conditions used (single hemiretinal stimulation of Cond. 5–12), LVF recognition levels decreased a mean 16.3%, compared to when full report was required, and RVF recognition levels increased a mean 23.0%. Analysis of all 16 report-stimulation conditions showed an overall RVF superiority of 22.5%, F (1, 4) = 39.74, $p < .01$, and, as with full report instructions, an overall temporal hemiretinal superiority, F (1, 4) = 28.19, $p < .01$.

Discussion

The present data support results normally found (unilateral RVF superiority, bilateral LVF superiority) for binocular and sometimes monocular viewing. Bilateral LVF superiority may confirm various reading direction hypotheses (White, 1969) and unilateral RVF superiority may confirm a directional and/or a cerebral dominance explanation. The present data, however, clearly show additional effects. Seeing a letter display with two hemiretinae is better than seeing it with one, and better with a temporal than a nasal hemiretina. These are the main findings of the present study, shown by two comparisons: monocular–binocular comparison and the triple hemiretinal comparisons.

Binocular viewing is generally superior to monocular viewing, whether unilateral or bilateral displays are being compared, and single temporal hemiretinal viewing results in recognition levels more similar to binocular levels than single nasal hemiretinal levels. In particular, the right eye's temporal hemiretina appears to be as efficient for recognition as is binocular viewing, even to the extent of reversing the typical unilateral RVF superiority by 7.0% (see Fig. 1). However, the superiority of temporal hemiretinal viewing is general, and we attribute the Eye × VF interactions, in both unilateral and bilateral displays, to more efficient temporal hemiretinal viewing. Comparisons of triple hemiretinal viewing with monocular and binocular viewing again demonstrate that two hemiretinae are better than one, and temporal better than nasal, the conclusion being made this time from conditions when only one visual field is seen by two hemiretinae. Adding another hemiretina to monocular viewing, or occluding one from binocular viewing, with greater effect if that hemiretina is temporal, allows the conclusion that these hemiretinal effects (two better than one, temporal better than nasal) are not specific to any particular combination of stimulation. For the experimental conditions and apparently visually symmetrical subjects employed in this study, at least, it appears possible to estimate binocular recognition scores from knowledge of monocular scores and information about hemiretinal effects. The present study consistently shows temporal hemiretinal superiority. To the extent that recognition of letters depends upon acuity, this result agrees with Markowitz and Weitzman's (1969) finding of superior gap detection of 1°45′ eccentric Landolt Cs viewed by temporal hemiretinae. But the present monocular and binocular bilateral results (Fig. 2) could hardly differ more from the results of reporting a spaced string of digits (see Crovitz & Lipscomb, 1963). In this latter study, *nasal* hemiretinal viewing was close to binocular viewing, and even though the stimulus displays differed from the present ones in size and type of material, and recognition was much lower, it is difficult to see why these differences should reverse the relative efficiency of nasal and temporal hemiretinae. Nasal superiority is also found in color rivalry (Crovitz & Lipscomb, 1963) and in reporting rivalrous digits (Bower & Haley, 1964; Sampson, 1969). Contralateral projections to the visual cortex (nasal hemiretinae) have been regarded as dominating ipsilateral projections in binocular co-operation and competition (Walls, 1953), but it would seem that in noncompetitive stimulation where

acuity is important in determining response accuracy, temporal superiority can be found.

The comparisons of number and type of hemiretinae made with the present data have always been within the basic unilateral or bilateral conditions, so that any explanations developed to account for unilateral–bilateral differences are not relevant to our examination of hemiretinal effects. Similarly, for left and right sides of the displays, letters were equally eccentric, and distance from fixation alone cannot explain the obtained hemiretinal effects. The magnitude of hemiretinal effects, however, is dwarfed by the effect of partial report. Partially reporting bilateral displays yields just as great RVF superiority as reporting unilateral displays (fully or "partially"), confirming Fitzgerald and Marshall's (1967) strong RVF superiority when partially reporting bilateral displays with binocular viewing. *Single* letter recognition or recall from unilateral (White, 1970) or bilateral displays (Smith & Ramunas, 1971; Winnick & Bruder, 1968) shows minor differences between total LVF and total RVF scores. Letter-position curves in these and related studies roughly follow either a symmetrical acuity gradient or a within-field report sequence effect, or a combination of the two, with, usually, peripheral positions superior or equal to penultimate positions. It would seem, therefore, that some type of serial processing of letter displays is involved in full and partial (half-field) report. Douglas's (1968) large RVF superiority (43.1%) when practiced *S*s reported RVF of a bilateral display first also points to the importance of report conditions. The partial report data of the present study, however, still displayed temporal hemiretinal superiority, and we would also expect the superiority of two hemiretinal viewing to be found regardless of report requirements.

REFERENCES

Barton, M. I., Goodglass, H., & Shai, A. Differential recognition of tachistoscopically presented English and Hebrew words in right and left visual fields. *Perceptual & Motor Skills*, 1965, **21**, 431–437.

Bower, T. G. R., & Haley, L. J. Temporal effects in binocular vision. *Psychonomic Science*, 1964, **1**, 409–410.

Corballis, M. C. Binocular interactions in letter recognition. *Australian Journal of Psychology*, 1964, **16**, 38–47.

Crovitz, H. F., & Daves, W. Tendencies to eye movements and perceptual accuracy. *Journal of Experimental Psychology*, 1962, **63**, 495–498.

Crovitz, H. F., & Lipscomb, D. B. Dominance of the temporal visual fields at a short duration of stimulation. *American Journal of Psychology*, 1963, **76**, 631–637.

Crovitz, H. F., & Schiffman, H. R. Visual field and letter span. *Journal of Experimental Psychology*, 1965, **70**, 218–223.

Douglas, B. Left-Right differences and order of report in tachistoscopic perception. Unpublished master's thesis, University of Auckland, 1968.

Fitzgerald, R. E., & Marshall, A. J. Left-right field differences with partial report of letters. *American Journal of Psychology*, 1967, **80**, 370–376.

Freeburne, C. M., & Goldman, R. D. Left-right differences in tachistoscopic recognition as a function of order of report, expectancy, and training. *Journal of Experimental Psychology*, 1969, **79**, 570–572.

Goodglass, H., & Barton, M. Handedness and differential perception of verbal stimuli in left and right visual fields. *Perceptual & Motor Skills*, 1963, **17**, 851–854.

Humphrey, M. E. Consistency of hand usage. *British Journal of Educational Psychology*, 1951, **21**, 214–225.

Kimura, D. The effect of letter position on recognition. *Canadian Journal of Psychology*, 1959, **13**, 1–10.

Koetting, J. F. Word recognition as a function of locus in the four lateral visual fields: The iota phenomenon. *American Journal of Optometry*, 1970, **47**, 55–66.

Markowitz, H., & Weitzman, D. O. Monocular recognition of letters and Landolt Cs in left and right visual hemifields. *Journal of Experimental Psychology*, 1969, **79**, 187–189.

McKeever, W. F., & Huling, M. D. Left-cerebral hemisphere superiority in tachistoscopic word-recognition performances. *Perceptual & Motor Skills*, 1970, **30**, 763–766.

Overton, W., & Wiener, M. Visual field position and word-recognition threshold. *Journal of Experimental Psychology*, 1966, **71**, 249–253.

Sampson, H. Recall of digits projected to temporal and nasal hemi-retinas. *Quarterly Journal of Experimental Psychology*, 1969, **21**, 39–42.

Smith, M. C., & Ramunas, S. Elimination of visual field effects by use of a single report technique: Evidence for order-of-report artifact. *Journal of Experimental Psychology*, 1971 **87**, 23–28.

Walls, G. L. The lateral geniculate nucleus and visual histophysiology. *University of California Publications in Physiology*, 1953, **9**, 1–100.

White, M. J. Laterality differences in perception: A review. *Psychological Bulletin*. 1969, **72**, 387–405.

White, M. J. Signal-detection analysis of laterality differences: Some preliminary data, free of recall and report-sequence characteristics. *Journal of Experimental Psychology*, 1970, **83**, 174–176.

Winnick, W. A., & Bruder, G. E. Signal detection approach to the study of retinal locus in tachistoscopic recognition. *Journal of Experimental Psychology*, 1968, **78**, 528–531.

(Received June 9, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 136–139

LEARNING TO CLUSTER [1]

CHARLES P. THOMPSON [2] AND DANIEL L. ROENKER

Kansas State University

In three experiments, *S*s received three trials on each of three successive categorized word lists. In all instances, the data show a dramatic learning-to-cluster effect with clustering on the first trial of the second list equivalent to clustering on the terminal trial of the first list. Intralist, trial-to-trial recall was, however, unaffected by the large changes in degree of clustering.

The term "learning to learn" has been applied to the observation that performance improves over successive tasks which are similar in nature. This improved performance presumably reflects the learning of the task requirements by *S*. But each learning task imposes multiple requirements (e.g., a typical free recall task involves pacing during presentation, written recall, etc.), and these requirements must be examined separately before it will be possible to determine what *S* is learning when he is "learning to learn."

When the task is the free recall of categorized word lists, one of the task requirements seems fairly obvious. Organization of the to-be-remembered material facilitates (e.g., Mandler, 1967; Sturges, Crawford, & Nelson, 1971) recall. Since organization aids recall and intralist organization increases with practice, it seems reasonable to hypothesize that learning-to-learn effects may reflect, at least in part, the increasing ability of *S* to organize the material for recall. That is, one of the task requirements is that *S* learn how to organize the material at hand, and once this has been accomplished, positive transfer effects may accrue to subsequent tasks of similar nature. In categorized word lists, increased organization is apparently reflected by increased clustering since clustering increases over trials and good clusterers recall more than poor clusterers (Thompson, Hamlin, & Roenker, 1971). Thus, the present experiments were analyzed to determine whether *S*s learn how to cluster categorized word lists.

Method

Experiment I

Subjects.—The *S*s were 60 summer school students at Kansas State University who were paid for their participation in the experiment. All *S*s were tested in groups of 10.

Materials.—Fourteen categories were selected from the Battig-Montague norms (1969). Twelve exemplars were selected from each category. Although individual items ranged in normative frequency from 1 to 362, the mean frequency per category was matched across all categories. A total of nine lists, composed of 4 categories each, were generated from these 14 categories.[3]

Experimental design.—A repeated-measures design was employed with all *S*s receiving three study-recall trials on each of three lists. To insure that the results were not list specific, *S*s were divided into three groups of 20 *S*s each with each group viewing a different set of three lists. No category was repeated in any three-list set. Each group was further subdivided such that half of the *S*s in each group saw one order of presentation and the other half saw a different order of presentation of the three lists.

Procedure.—The words were presented at a 2-sec. rate with a Kodak Carousel projector during the study portion of each trial. A different random order of presentation was used for each trial. The

[1] This research was supported in part by a predoctoral fellowship in experimental psychology to the second author from National Institute of Mental Health Experimental Training Grant MH-08359-06 and in part by a grant to the first author from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed herein, however, do not necessarily reflect the position or policy of the United States Office of Education, and no official endorsement by the United States Office of Education should be inferred.

[2] Requests for reprints should be sent to Charles P. Thompson, Department of Psychology, Kansas State University, Manhattan, Kansas 66502.

[3] The unusual combination of categories within lists stems from the fact that the studies reported here originally included a fourth list, which repeated categories from previous lists in an attempt to produce proactive interference. Since the data from those lists represent a combination of both learning-to-learn and proactive interference effects, they were excluded from the present analysis.

Ss, during the presentation of each word, identified the category membership of the word by writing the category initial in an answer booklet. This procedure was employed to assure that Ss attended to each word.

Following the final word in the list, a three-digit number was presented. The Ss filled a 30-sec. interval between presentation and recall by counting backwards by three's from that number at a 1-sec. rate. This filled interval presumably eliminated the short-term memory component in recall (Glanzer & Cunitz, 1966).

Two minutes were allowed for recall on each trial. The Ss were permitted to recall the words in any order that occurred to them. Each study and recall protocol was recorded on a separate sheet in the answer booklet. A 30-sec. unfilled interval was used between trials and a 3-min. interval between lists.

Experiment II

Subjects.—The Ss were 60 introductory psychology students at Kansas State University who participated in the experiment in partial fulfillment of course requirements. All Ss were tested in groups of 10.

Materials, design, and procedure.—The materials, design, and procedure were identical to those used in Experiment I.

Experiment III

Subjects.—The Ss were 30 summer school students at Kansas State University who were paid for their participation in the experiment. All Ss were tested in groups of 10.

Materials.—Twenty-six categories were selected from the Battig-Montague norms (1969). Each category consisted of seven exemplars. Exemplars were chosen such that the mean frequency per category was matched across all categories. Individual items ranged in frequency from 10 to 426. Further, the Thorndike-Lorge mean frequency of occurrence in the English language per category was matched across all categories with the range of individual item frequencies being 1 to AA. A total of seven lists of 6 categories each were generated from these 26 categories.

Experimental design.—A repeated-measures design was employed with all Ss receiving three study-recall trials on each of three lists. To insure that the results were not list specific, Ss were divided into three groups with each group viewing a different set of three lists although, in this case, the same first list was used for all three groups. No category was repeated in any three-list set.

Procedure.—The procedure was identical to that employed in Exp. I and II.

RESULTS

Within each experiment, recall and clustering scores were computed for each S for each trial of each list. The index of clustering used, ARC, fixes chance and perfect clustering at 0 and 1, respectively. Thus, the measure represents the relative amount of clustering between chance and perfect performance. A detailed description of the ARC measure may be found in Roenker, Thompson, and Brown (1971).

Experiment I

Clustering.—The mean ARC score for each trial for each list is presented in the upper portion of Table 1. An analysis of variance performed on the clustering data[4] demonstrated that intralist organization increased over trials, $F\ (2, 116) = 29.97$, $p < .001$. More importantly, Ss improved their clustering performance over successive lists, $F\ (2, 116) = 55.95$, $p < .001$.

[4] During initial data analysis, a single S's protocol was inadvertently destroyed. Thus, the analyses reported are based on an N of 59 instead of 60.

TABLE 1

MEAN ADJUSTED RATIO OF CLUSTERING (ARC) AND RECALL SCORES FOR EACH TRIAL OF EACH LIST FOR EXPERIMENTS I AND II

Exp.	List 1			List 2			List 3		
	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
I									
ARC	.49	.65	.71	.70	.83	.82	.75	.80	.87
Recall	14.61	21.00	23.08	16.59	23.24	26.14	16.78	22.69	25.61
II									
ARC	.41	.67	.75	.63	.80	.85	.73	.85	.81
Recall	14.83	20.93	24.42	16.73	23.48	25.97	17.48	23.50	25.88

However, as shown by the significant Lists × Trials interaction together with a subsequent Newman-Keuls test, the increase in clustering performance across lists manifests itself in a dramatic increase between List 1 and List 2, with no reliable improvement thereafter, $F(4, 232) = 2.90, p < .05$.

Recall.—Mean recall for each trial for each list is also presented in the upper portion of Table 1. Recall performance increased over trials, $F(2, 116) = 282.99$, $p < .001$, and over lists, $F(2, 116) = 17.67$, $p < .001$. The Lists × Trials interaction was not statistically significant.

Experiment II

Clustering.—The results were virtually identical to Exp. I, with clustering performance increasing reliably across lists, $F(2, 118) = 20.54, p < .001$, and intralist organization increasing across trials, $F(2, 118) = 54.21, p < .001$. The dramatic increase from List 1 to List 2 in clustering performance once again obtained and is reflected in the significant Lists × Trials interaction, $F(4, 236) = 6.68, p < .001$. In this experiment, the Newman-Keuls test on the interaction means showed a reliable increase in Trial 1 clustering from List 2 to List 3, with no comparable increase for Trials 2 and 3. The trial-by-trial clustering data for Exp. II are shown in the lower portion of Table 1.

Recall.—As in Exp. I, recall increased over trials, $F(2, 118) = 302.88, p < .001$, and over lists, $F(2, 118) = 15.97, p < .001$. The Lists × Trials interaction was not reliable. The mean recall data are also presented in the lower portion of Table 1.

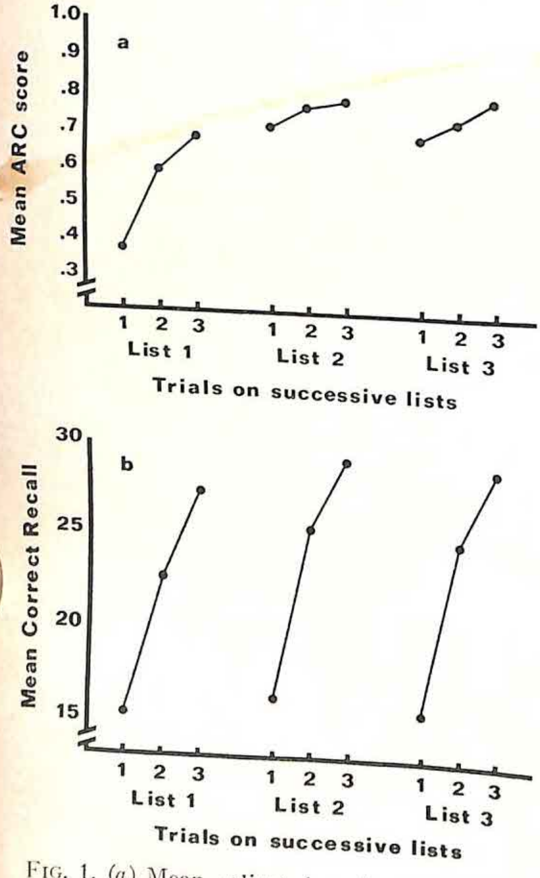

FIG. 1. (*a*) Mean adjusted ratio of clustering (ARC) scores for each trial for each of three lists or Exp. III, and (*b*) mean correct recall for each trial for each of three lists for Exp. III.

Experiment III

Clustering.—Once again, clustering increased reliably across lists, $F(2, 58) = 16.19, p < .001$, and trials, $F(2, 58) = 18.42, p < .001$. As before, the Lists × Trials interaction was significant, $F(4, 116) = 4.58, p < .01$. These data are presented in the upper portion of Fig. 1. A Newman-Keuls test on the interaction means showed a significant increase in clustering from List 1 to List 2, with no reliable change thereafter.

Recall.—For comparative purposes, the recall means are presented in the lower portion of Fig. 1. As can be seen, recall improved over trials, F $(2, 58) = 360.04$, $p < .001$, and over lists, $F(2, 58) = 5.09$, $p < .01$. As before, the Lists × Trials interaction was not statistically significant.

DISCUSSION

The data unambiguously demonstrate that *S*s learn to cluster over successive categorized word lists. The most rapid changes in clustering performance take place between List 1 and List 2, with essentially no change thereafter (see Fig. 1). The change in clustering performance over the first two lists is dramatic, with first-trial performance on List 2 equivalent to terminal-trial performance on List 1.

In short, the data show that learning to cluster occurs and is essentially complete after multiple trials on a single categorized word list.

In contrast to the clear learning-to-cluster effect, the relationship between the clustering and recall data does not suggest an unambiguous interpretation. Although both clustering and recall increase over lists, the higher degree of clustering in the second and third lists is not accompanied by a corresponding improvement in intralist recall—i.e., trial-to-trial increments in recall are unaffected by degree of clustering. This pattern of results suggests two alternative interpretations. The most obvious possibility is that clustering and recall performance are unrelated. However, it is equally possible that within limits, improved clustering results in improved recall but that the effective improvement in clustering is complete after the first trial. Under this view, the relatively poor recall performance on all trials of List 1 may be attributed entirely to the poor clustering performance on Trial 1 of that list. While these two alternatives can not be distinguished on the basis of the data from the present experiments, perhaps the most serious consideration should be given to the latter alternative since other evidence (e.g., Mandler, 1967) suggests that organization aids recall.

REFERENCES

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut Category Norms. *Journal of Experimental Psychology*, 1969, **80** (3, Pt. 2).

Glanzer, M., & Cunitz, A. R. Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 351–360.

Mandler, G. Organization and memory. In K. W. Spence & Janet T. Spence (Eds.), *The psychology of learning and motivation*. Vol. 1. New York: Academic Press, 1967.

Roenker, D. L., Thompson, C. P., & Brown, S. C. A comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, 1971, **76**, 45–48.

Sturges, P. T., Crawford, C. B., & Nelson, A. K. The effects of instructions upon free recall and subjective organization of abstract and concrete words. Paper presented at the meeting of the Midwestern Psychological Association, Detroit, May 1971.

Thompson, C. P., Hamlin, V. J., & Roenker, D. L. A comment on the role of clustering in free recall. *Journal of Experimental Psychology*, 1971, in press.

(Received June 14, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 140–153

RESPONSE MECHANISMS IN DETECTION EXPERIMENTS [1]

WILLARD LARKIN [2]

University of Maryland

Listeners rated the subjective loudness of tones presented in noise and made decisions as to the presence or absence of Signals. Loudness ratings and detection decisions were studied first in separate experiments and then in experiments which required both responses to be made on each trial. The loudness judgments were stable across sessions and Signal probability conditions, and the two responses could be made on the same trials with no discernible interference. The relation between the two responses was compared with predictions from psychophysical models. While the data support the view that detection responses are composed of sensory and decision stages, they are inconsistent with several traditional models. A variable criterion model is proposed which gives a good account of the detection data.

In contemporary theory, a two-stage mechanism underlies perceptual responses: the stimulus is first transformed into its sensory representation and then *S* determines his response in accord with the decision-making aspects of his task. Ordinarily, the two response stages are not studied in isolation, and their separate properties must be inferred from data that reflect them both. But the sequential, noninteractive form of the two-stage theory suggests that an empirical decomposition should be possible. This paper reports experiments designed to separate the response stages in an auditory detection task, in order to permit an analysis of the way detection responses relate to their sensory antecedents. Several models for the decision stage are used to focus the analysis.

In the usual "Yes/No" detection paradigm, *S* must decide on each trial whether a "Signal" had occurred or not. The present *S*s had to do more. In the initial experiments, they rated the subjective loudness of an auditory stimulus, without indicating whether it represented a Signal or a Nonsignal. Later, the Signal/Nonsignal decision was appended to the rating task, creating a *dual-response* trial in which *S*s were to assume, first, an unbiased attitude for reporting sensory experience, and, second, a decision-making attitude enabling a detection judgment to be made. Each kind of response was investigated separately, but the dual-response trials provided the primary data for the evaluation of decision models.

The first objective, then, was to produce loudness judgments closely related to the energy dimension of the stimuli, but invariant with respect to decision parameters of the detection task. These judgments could then provide an index of the sensory background for each detection response. The next step was to introduce variation in one of the decision parameters—the Signal probability—and to explore the resulting pattern of detection responses on dual-response trials. Whether the data validly may be compared with two-stage theory depends, of course, upon whether the loudness ratings are stable indicators of *S*'s sensory input. Models for the decision stage are described in the section on Theory, with details in the Appendix. Principal attention is given to a simple random walk mechanism for a variable response criterion; this is compared with learning models for fluctuating response bias, and with other traditional theories.

[1] This research was supported in part by a grant-in-aid from the Society of the Sigma Xi by United States Public Health Service Grant NB 05636 and by the University of Maryland Computer Science Center.

[2] Some of this work was reported in a PhD dissertation submitted to the Graduate College of the University of Illinois. The author is grateful to Gordon Z. Greenberg and Harold W. Hake for their generous attention to this project and to Robert B. Ash, William H. Batchelder, and Ledyard R. Tucker for their helpful comments and suggestions.

Method

Four *S*s, in acoustic isolation, listened to the same binaural stimuli. Each used a set of Permoflux PDR-10 headphones and made responses on a panel fitted with pushbuttons and indicator lamps. The *S*s were tested in daily 2-hr. sessions divided into blocks of 60–85 trials, with a rest period midway in each session.

Each trial began with a 1.3-sec. warning light, and a 1-kHz. Standard Tone of 250-msec. duration, presented in a continuous background of 20-kHz. bandwidth Gaussian noise. The noise was about 55 db. above threshold (1.598×10^{-3} V. *rms* at the headphones). The Standard Tone (1.610×10^{-3} V. *rms*., for continuous input at the headphones) nearly always could be detected.

The Standard Tone and the warning light were turned off simultaneously, and 750 msec. later, a 1-kHz. Test Tone was presented for 100 msec., together with an indicator light. The Test Tone occurred at one of four randomly selected levels, denoted T_1, T_2, T_3, and T_4. The highest, T_1, was 2 db. below the Standard Tone, and the others were spaced 2 db. lower in succession. These four levels spanned the range of imperfect detectability: in Yes/No tests with other *S*s, correct responses were given on about 95% of T_1 trials, and about 55% of T_4 trials. (Chance performance is 50%.)

In conditions with Signal feedback information (provided by indicator lamps), the Signal, denoted E_1, was operationally equivalent to the presentation of either of the higher level stimuli, T_1 or T_2, and a Nonsignal, denoted E_0, was a presentation of either of the lower level stimuli, T_3 or T_4. The *S*s had no knowledge of this correspondence with feedback events nor of the limitation of stimuli to four intensity levels. The use of four levels is a departure from conventional Yes/No procedure, in which E_1 is a single nonzero intensity and E_0 is a "blank" trial. (See Green & Swets, 1966, for examples of the usual procedure.) This change increases the range of the loudness dimension, which permits a closer analysis of rating stability, and of *S*'s decision pattern.

A 100-msec. blank interval followed the Test Tone, and *S* could then press a loudness rating pushbutton or a Yes/No pushbutton, depending on the experiment. In the dual-response experiments, both responses were made in succession. The time alloted for either response was 3 sec. The four loudness categories are denoted R_1, R_2, R_3, and R_4, in the order "loud" to "soft," and the Yes/No alternatives are denoted A_1/A_0, respectively.

The Signal probability, $P(E_1)$, was controlled by sampling a punched tape of random digits. On each trial, E_1 or E_0 was selected first, and one of the two corresponding Test Tones was selected next, with equal probability. During any block of trials, and during sets of three or more blocks, $P(E_1)$ was constant. Variations in $P(E_1)$ were introduced to induce changes in *S*s' decision behavior and to test loudness rating stability under shifts in the stimulus distribution. The procedure began with sessions devoted only to loudness ratings, and with $P(E_1) = .50$. This *standard rating condition* was repeated as a "calibration condition" in the first three blocks of every session. In later sessions, $P(E_1)$ was set at .25, .40, or .70. Signal feedback information was added in some of these sessions, before *S*s were required to make detection decisions. Finally, *S*s made both loudness ratings and detection decisions on each trial, with Signal feedback information.

Instructions.—The physical nature of the stimuli was described for *S*s in some detail, except that no information was given as to the actual levels of the Test Tone. Fluctuations in its loudness were explained as a joint consequence of the effects of noise and of inherent random variations in the Test Tone itself. For the rating experiments, *S*s were told:

> We are interested in how well people can sort out sounds of different loudnesses in situations like this, where the background partly determines whether the sound appears loud or soft. Your task is to make judgments of the loudness of the Test Tone on each trial. Ask yourself: how strongly did the tone come through the noise; and then give a rating with one of these four pushbuttons . . . . Because of the noise, the Standard Tone will not seem equally loud on all trials. Thus, try to develop your own idea of a typical Standard Tone, based upon hearing it several times. Make judgments of the Test Tone in relation to this "ideal" Standard Tone.

Whenever changes were made in $P(E_1)$, instructions were given to promote consistency:

> In the next condition, the average loudness of the Test Tone will be higher (lower) than in the last condition. The difference will not be great, but it will be noticeable. Undoubtedly, you will use the louder (softer) rating categories more often, but it is very important that your response be absolutely consistent with the previous condition. In other words, your rating criteria should be the same.

Whenever Signal feedback was given, the change in "average loudness" was described as resulting from a change in $P(E_1)$, but only the direction of change, not the probability value, was announced. Signal events were defined by the feedback lamps, but were described as related to the intensity of the Test Tone:

> When the Signal occurs, the Test Tone is slightly louder than it is when the Signal does not occur. Thus, the loudness of the Test Tone tells you about the presence or absence of a Signal. The difference in loudness between Signal and Nonsignal trials is quite small.

For dual-response trials, *S*s were asked to make the loudness rating *before* considering the Yes/No decision. The goal of the Yes/No response was to "make as many correct decisions as possible."

Theory.—In the procedure just described, the loudness ratings simulate the sensory stage in the de-

tection process. To test predictions about the decision stage, these ratings must exhibit the stability, and the independence from decision-making factors, that most theories postulate for S's sensory representation. Ordinarily, psychophysical judgments do not have this stability (Garner, 1954; Parducci, 1963), although they may be reliable if experimental conditions are constant (Bell & Nixon, 1971; Green, 1964). In detection tasks, one source of disturbance may be that Ss commonly are instructed to base their ratings on subjective "confidence" (Clarke, 1960; Schulman & Greenberg, 1970; Swets, Tanner, & Birdsall, 1961). If Ss interpret this instruction to mean that the Signal probability should be taken into account, their responses may not faithfully reflect internal sensory events.

One might argue that just as in perceptual adaptation to suprathreshold stimulation (Helson, 1964), $P(E_1)$ may, through its effect on the stimulus distribution, induce real changes in the sensory effect of any one stimulus. Nonetheless, it seems important to determine whether judgmental instability can be avoided, first because of the prevailing notion in detection theory that presentation probabilities do not affect Ss' sensations and, second, because bias-free ratings would permit a closer analysis of decision patterns in dual-response trials. Although the response models examined in this paper represent the decision stage differently, they share the assumption that S's sensory stage does not depend on $P(E_1)$.

The traditional view of the way Ss generate detection decisions is that they use a covert criterion, which may be a sensory threshold (Blackwell, 1953; Boumann, 1955; von Bekesy, 1960) or a response threshold determined by the probabilities, costs, and values of response outcomes (Green, 1960; Tanner & Swets, 1954). The present experiments do not examine the sensory threshold hypothesis, but they provide a test of the likelihood-ratio criterion of signal-detectability theory. If Ss heed the instruction to maximize correct A_i decisions, their decision criteria should be proportional to $P(E_1)/P(E_0)$, with a proportionality constant, V, which reflects the imbalance, if any, in subjective costs and values. (See Green & Swets, 1966, for a complete presentation of the likelihood-ratio hypothesis.) If S's decision criterion is determined in different $P(E_1)$ conditions, V can be estimated and checked for invariance. With loudness ratings as an index of the sensory effect on each trial, the determination of decision criteria is necessarily imprecise. Nevertheless, the pattern of estimates across $P(E_1)$ conditions may be sufficient to judge the likelihood-ratio prediction.

Although the likelihood ratio may be better suited than a sensory threshold as a basis for detection decisions (Swets, 1961), these ideas share a faith in the optimality of the response process. Both theories posit a criterion that is fixed by experimental conditions. It is likely that whatever covert criterion S may use, it may change from trial to trial. Data on sequential aspects of detection responses (e.g., Carterette, Friedman, & Wyman, 1966; McGill, 1957; Shipley, 1961) are not inconsistent with this possibility.

A simple model for changing criteria is outlined in the Appendix. The basic assumption is that Ss shift their response criteria after (some) error trials in a way that renders the same error less likely on future trials. A version of this adaptive mechanism was proposed by Kac (1962) as an extension of signal-detectability theory, in which stimulus effects have a Thurstonian representation on a continuous scale. The model described in the Appendix substitutes a discrete scale of sensory effects (which may be interpreted as a partition of a continuous scale). This change provides a natural correspondence with the loudness rating experiments and makes the resulting random walk more accessible to analysis than Kac's original proposal.[3]

The simplest version of the random walk assumes that S sets a decision criterion between adjacent steps on the scale of sensory magnitude: if the stimulus leads to a sensory effect above the criterion, A_1 is given; if the effect is below the criterion, A_0 is given. When responses are in error, the criterion moves one step in a direction that makes the incorrect response less probable, and these adjustments generate a distribution of criterion positions. The parameters of the distribution will depend on the presentation probabilities for the various stimuli and, thus, on $P(E_1)$. The model can be evaluated by estimating the criterion distributions for different $P(E_1)$ conditions, and using these estimates to predict the pattern of A_1/A_0 decisions. For the basic version of the model, as assumed in this paper, the criterion distribution in any $P(E_1)$ condition can be determined from any other $P(E_1)$ condition; hence the decision patterns in several conditions can be predicted jointly, from a single independent estimate of one criterion distribution. When loudness ratings are identified with steps on the sensory scale, a single rating experiment provides this estimate. For the present analysis, the *standard rating condition* supplies the criterion distribution used for all predictions.

A fundamentally different view of the response process is to regard detection decisions as inherently probabilistic (or uncertain), rather than as determined by a fixed or variable criterion. This characterization is in the spirit of the Brunswikian model of perception (Hammond, 1966), which attaches importance to the imperfect predictive validity of sensory (or proximal) "cues" to the stimulus. If S makes detection decisions according to the predictive validities of his sensory states, his response pattern may share some of the statistical aspects of responses in multiple-cue prediction tasks. He may, for example, maintain separate response

[3] Whether sensory effects are best represented as continuous or discrete is an unsolved question. The position taken here is that in the absence of convincing evidence for either view, the nature of S's task may determine which is appropriate. For a discussion of this question in relation to rating experiments, see Krantz (1969).

biases for different levels of sensory input, as Taylor's (1966) *Ss* did for cues with different levels of validity. If levels of sensory input serve *S* as cues to a prediction, the response probabilities at each level may tend to match the corresponding Signal probabilities, as in Lee's (1963) simulation of a signal-detection experiment. Generally, one would expect that whenever a sensory state is an imperfect indicator of the Signal, the detection response will be uncertain. To make the expectation precise, it is necessary to postulate a specific model for probabilistic responding, but the general hypothesis can be assessed simply by inspecting the pattern of response probability in the present experiments.

Models for probabilistic responding in detection experiments often are cast in the framework of stochastic learning theory (Atkinson, 1963; Atkinson & Kinchla, 1965; Bush, Luce, & Rose, 1964; Luce, 1963b). While these models differ in detail, their central postulate is the same: that some of *S*'s responses result from a bias which is learned in the course of the experiment. Because the learning depends on feedback events, the bias fluctuates from trial to trial. The overall effect may be quite similar to that of a shifting response criterion, and a detailed sequential analysis may be required to distinguish a learning mechanism from a criterion that undergoes a random walk.

To focus the analysis of detection data, it is useful to compare the random walk description with some specific learning assumptions. Two such models, based upon the linear operator scheme of Bush and Mosteller (1955), are outlined in the Appendix. Both place rather strong constraints on the learning process, and both thereby make predictions quite different from the random walk model. The first requires a single linear operator, which is applied independently for each of *S*'s sensory states. Denoting the sensory states D_k, this *independent transitions* model allows changes in the response conditional probability $P(A_1{:}D_k)$ only on trials when *S* is in State D_k. Thus, fluctuations in response tendencies occur independently for each discrete sensory level. The second model permits complete sensory generalization, which means that $P(A_1{:}D_k)$ can fluctuate simultaneously in all states, but with separate rates for each. This model, therefore, has *independent operators*, which reflect the differential importance of outcome feedback for the various levels of sensory input. The experiments provide a test of both models' parameter invariance, as well as their basic postulates about the effects of feedback.

The question of sensory generalization is rarely dealt with in learning models for psychophysics because it is often assumed that *S*'s response uncertainty is confined to a single sensory state. (Two exceptions are the models proposed by Bush, Luce, & Rose, 1963, and by Schoeffler, 1965). When response uncertainty is not so confined, as in the models just described, it may be unrealistic to suppose either that *Ss* fail to generalize feedback information from one state to another (the independent transitions model) or that the generalization is without regard for the natural ordering of sensory effects (the independent operator model). A reasonable alternative is a mechanism that would permit *directional sensory generalization.* To give an example, assume that if state D_j occurs on Trial n, then with respect to $P_n(A_1{:}D_k)$,

$$P_{n+1}(A_1{:}D_k)\begin{cases}\text{is increased if} & E_{1,n} \text{ and } j \le k.\\ \text{is unchanged if} & E_{1,n} \text{ and } j > k \text{ or}\\ & E_{0,n} \text{ and } j < k.\\ \text{is decreased if} & E_{0,n} \text{ and } j \ge k.\end{cases}$$

In this scheme, each Signal occurrence (E_1) increases the probability of Yes (A_1) if the sensory state reflects a higher magnitude than the state on Trial n; and each Signal nonoccurrence decreases the probability of A_1 if the sensory state reflects a lower magnitude. A similar idea was advanced by Schoeffler (1965) for a continuous sensory scale. One consequence of such a mechanism, but not of the two learning models previously discussed, is that $P_{n+1}(A_1{:}D_k)$ will be a nonincreasing function of the state index, j, of Trial n: as subjective loudness increases, the A_1 response becomes less likely on the succeeding trial. This is a property that the random walk model shares. Any attempt to distinguish it from the generalization learning scheme will, therefore, require more detailed analysis of data.

Results and Discussion

Loudness rating stability.—We first examine rating stability over sessions. To do this, data have been grouped into three experimental stages: five initial sessions run before information about Signal events was introduced (Stage A: 1,071 trials); four intermediate sessions (Stage B: 698 trials); and five later sessions (Stage C: 976 trials). Data for Stages B and C came from initial blocks of each session: these were "rating calibration" blocks prior to dual-response trials. In all three stages, the stimulus distribution was uniform, with $P(E_1) = .50$, and so the data are replications of the *standard rating condition.*

Tables 1 and 2 present measures of stability for each *S*. Unfortunately, no single statistic is appropriate for this purpose, and the data are too numerous to describe in detail. The first statistic given in Table 1 is $U(R)$, the uncertainty measure for the response distributions, which may be viewed as an index of response variance, and, to some extent, of the distributional configuration. Across the three stages, this index varies less than 2%

TABLE 1

INFORMATION ANALYSIS FOR LOUDNESS RATINGS

S	Data stage		
	A	B	C
1	1.810 .119	1.795 .119	1.788 .084
2	1.986 .106	1.981 .111	1.962 .094
3	1.748 .112	1.782 .097	1.785 .098
4	1.997 .094	1.902 .086	1.936 .107

Note.—The upper rows list the response uncertainty, $U(R)$, for each S; the lower rows list $U(R:T)/U(R)$, where $U(R:T)$ is the information transmitted.

for the first three Ss, and about 5% for S_4. The information transmitted (in the sense of absolute judgments) was also computed and is given in the second row of Table 1 as a (dimensionless) proportion of response uncertainty. On the average, Ss transmitted about .18 to .20 bits per response, and the standard deviation of this value, computed over stages, amounted to about 10% of the mean and to less than 1% of the total response uncertainty.

More direct information about the stability of $P(R_k:T_j)$ estimates is given in Table 2. The range of these estimates was determined over the three data stages for each of the 16 (k, j) pairs: the mean of these ranges is reported for each S. Again, S_4 appears the more inconsistent. A gradual shift in the skewness of his response distribution was discernible as one

TABLE 2

CONSISTENCY OF LOUDNESS RATINGS ACROSS DATA STAGES AND SIGNAL PROBABILITIES

S	Standard ratings: Stages A, B, C	$P(E_1)$ = .25, .50,[a] .70
1	.069	.049
2	.058	.099
3	.048	.067
4	.102	.088

Note.—Each entry is the mean range of $P(R_k:T_j)$ estimates for all pairs, (k,j), from three independent sets of data.
[a] Data from Stage C.

factor contributing to inconsistency, and, because of this, the analyses reported later in this paper are based mainly upon data from Stage C.

Table 2 also lists the mean range scores across three Signal probability conditions: $P(E_1)$ = .50 (Stage C data); .25 (575 trials); and .70 (620 trials). Data for the two latter conditions came from dual-response trials. When compared, the two columns of Table 2 provide an assessment of rating stability under changes in the distribution of stimuli, and with a detection decision required after each rating. Some increase in variability might be expected simply because the overall sample sizes are somewhat smaller for these conditions than for the standard ratings. But no general increase is evident. Changes in $P(E_1)$ generate differences in the rating probabilities that are about the same size as differences among replications of the standard ratings. For both sets of data, the expected largest difference among three estimates of $P(R_k:T_j)$ is between 5% and 10%.

If loudness ratings are to be used as sensory cues to the Signal, a fundamental requirement is that the posterior probabilities of the Signal, $P(E_1:R_k)$, remain constant across $P(E_1)$ conditions. Of course, this predictive accuracy is ensured if the rating probabilities $P(R_k:T_j)$ are constants, for then $P(E_1:R_k)$, which is given by the Bayesian calculation,

$$\frac{P(R_k:E_1)P(E_1)}{P(R_k:E_1)P(E_1) + P(R_k:E_0)P(E_0)},$$

will be determined only by the Signal probability. Moreover, some of the variability shown in Table 2 may not survive the Bayesian calculation because stimulus pairs are combined into Signal (E_1) and Nonsignal (E_0) events. Nevertheless, it is informative to check the consistency of $P(E_1:R_k)$ estimates and to examine whether they are experimentally distinct for different $P(E_1)$ conditions. One check on predictive consistency is illustrated in Fig. 1, which plots $P(E_1:R_k)$ estimated from ratings in the highest (.70) and lowest (.25) $P(E_1)$ conditions, using 625

and 430 trials, respectively. These experiments were run after *S*s had been introduced to the concept of a Signal, and with the feedback lamps operating, but without the requirement for detection decisions. (Experiments with $P(E_1) = .40$, run under these conditions, are not shown.) The solid lines in Fig. 1 represent the values predicted from Stage C ratings via the Bayesian formula. The discrepancies show no obvious or systematic pattern and are in the same range as discrepancies among similar estimates (not shown) for Stages A and B.

A further test of predictive stability is the comparison, in Fig. 2, of the same calculations from Stage C data with ratings from dual-response trials. Data are given for three $P(E_1)$ conditions (450 trials with $P(E_1) = .40$; other sample sizes are as given for Table 2). This comparison bears on the question of response interaction. If, contrary to instructions, *S* first covertly decided his detection response and then picked a rating to be consistent with it, the posterior probability functions might show some systematic distortion. Although such behavior cannot be prevented in

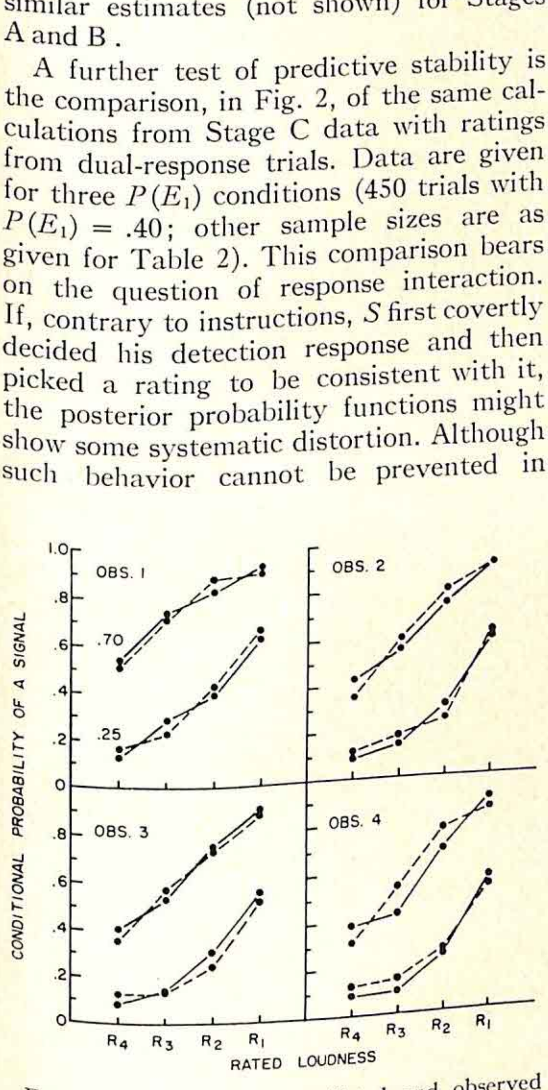

FIG. 1. Comparison of predicted and observed values of $P(E_1:R_k)$ for two $P(E_1)$ conditions. (Predicted values are shown by solid lines; observed values by dashed lines. Upper plots: $P(E_1) = .70$; lower plots: $P(E_1) = .25$. All predictions are from Stage C ratings, with $P(E_1) = .50$.)

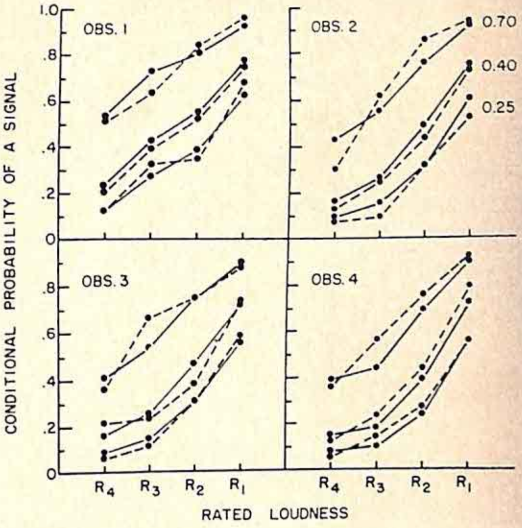

FIG. 2. Comparison of predicted values of $P(E_1:R_k)$ with ratings from dual-response trials. (Dashed lines indicate data; solid lines are predictions from Stage C ratings. Unlabeled numbers indicate the three Signal probability conditions.)

these experiments, the consistency in Fig. 2, together with the data already discussed, suggests that this did not happen.

The foregoing results demonstate that within the usual limits of a detection situation, *S*s are able to make loudness judgments with a degree of consistency not often found in detection responses or ratings. With explicit instructions to maintain constant subjective criteria, and given sufficient practice, these judgments can be relatively stable indicators of stimulus energy, without yielding noticeably to changes in the distribution of energies presented. Whether detection *S*s normally do this covertly is a question these experiments do not answer, but the data suggest that contemporary theory may not be wrong to characterize response bias as something appended to the representation of sensory effects rather than as something which alters the representation.

Evaluation of decision models.—The separation among posterior probability functions (see Fig. 2) seems sufficient, at least for the extreme conditions, to require major bias shifts in the detection decisions. The extent of this effect on psychometric functions can be inspected in Fig. 3. The

solid line plots the probability of a "Yes" (A_1) response against the four stimulus levels, Tj. These data from dual-response trials show the expected shift as the Signal probability changes from .25 to .70. Also shown are data from single-response detection trials (dashed lines) run before the dual-response experiments to check the consistency of detection decisions. While the functions match in most cases, there is a slope discrepancy at stimulus T_3, as indicated by the 95% confidence bands. The reason for the discrepancy is unknown. Because there are fewer observations where the discrepancy is greatest (60 T_3 trials in the single-response .70 condition, vs. 190 trials in the .25 condition), the slope differences may reflect differences in the amount of decision-making practice.

We next examine the relation between Yes/No decisions and loudness categories, and the way in which this relation depends on the asymmetry in the stimulus presentation schedule. These data are plotted in Fig. 4 for three $P(E_1)$ values: .25 (575 trials); .40 (448 trials); and .70 (620 trials). Although the results are less uniform across Ss than previous data involving only one response, certain patterns emerge. First, there is no fixed dependence of the detection decision on rated loudness: whether A_1 or A_0 is chosen depends strongly on $P(E_1)$. Second, fewer than half of the estimates of $P(A_1:R_k)$ are substantially different from zero or one. Thus, while the response pattern shifts with $P(E_1)$, it is always anchored at an extreme probability, with appreciable response uncertainty at only one or two points. Most of the inter-S variability is concentrated at these points.

To the extent that loudness ratings simulate sensory states, these results support the notion that probabilitistic responding may be confined to a narrow sensory region, a region somewhat smaller than the range of imperfect detectability used in these experiments. While this finding agrees with many discrete-state models, it is contrary to some, in that no single region (loudness category) absorbs all the response uncertainty across $P(E_1)$ conditions: Ss adjust to changes in $P(E_1)$ by shifting their decision uncertainty from one region to another.

The dependence of detection decisions on $P(E_1)$ is predictable from several points of view discussed earlier in this paper. According to one view, $P(A_1:R_k)$ should closely match $P(E_1:R_k)$. According to another—the signal-detectability hypothe-

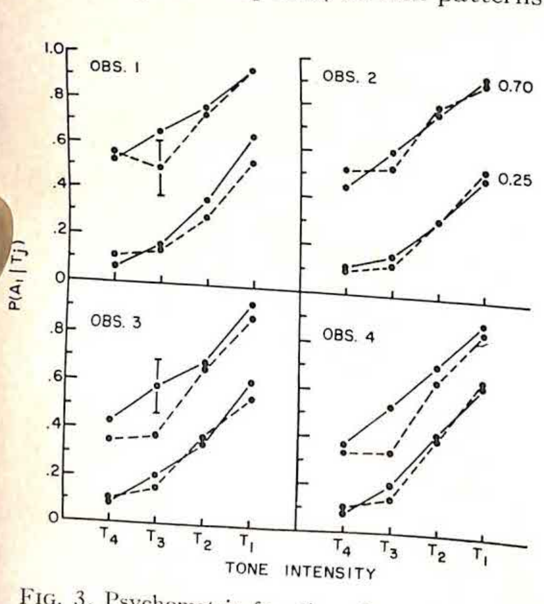

FIG. 3. Psychometric functions from single- and dual-response experiments. (Solid line: data from the dual-response condition. Dashed line: data from the single-response condition. Unlabeled numbers indicate the Signal probability.)

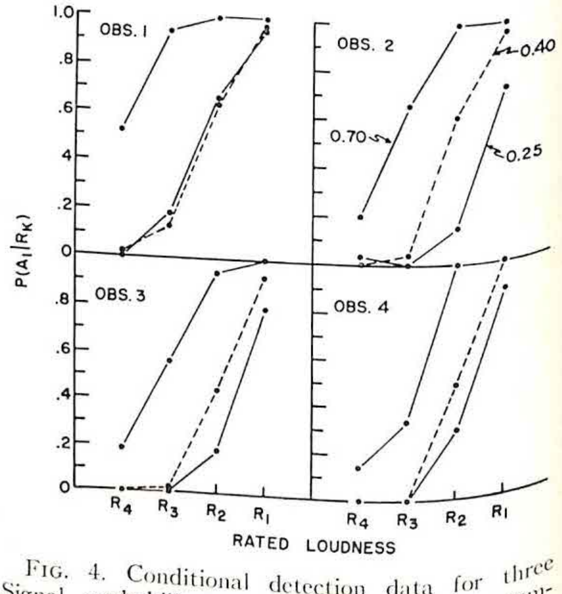

FIG. 4. Conditional detection data for three Signal probability conditions. (Unlabeled numbers indicate $P(E_1)$.)

sis—$P(A_1:R_k)$ should change sharply at a "criterion" value of $P(E_1:R_k)$. The observed relation between these two conditional probabilities, plotted as triangular points in Fig. 5, 6, and 7, confirms neither of these predictions. The plots are reminiscent of data from binary prediction studies, in which *S*s generally do not match event probabilities, but produce response functions somewhat steeper than the diagonal (Luce & Suppes, 1965; pp. 390–395).[4] With no evidence for sharp response cirteria, the likelihood-ratio hypothesis of signal-detectability theory could be checked only approximately, by estimating a criterion locus for each $P(E_1)$ value. To do this, smooth curves were drawn through each plot, and the crossover points (corresponding to $P(A_i:R_k) = .50$) were determined graphically. For no *S* were these points expressible as $V \cdot P(E_1)/P(E_0)$, without allowing threefold variation in V, correlated with $P(E_1)$. In general, *S*s undercompensated for shifts in Signal probability and thus appeared conservative with respect to the optimal criteria.

Two remaining points of view are the stochastic learning hypothesis and the variable criterion (random walk) hypothesis. We first examine two specific learning models for parameter invariance. For the independent transitions model, asymptotic values of $P(A_i:R_k)$ depend on a parameter, ϕ, which, for any fixed $P(E_1)$, should be constant across rating categories. The independent operator model has four parameters, ϕ_1, ϕ_2, ϕ_3, and ϕ_4, which correspond to the four rating categories; these should be constant across $P(E_1)$ conditions. Estimation equations for both models are given in the Appendix, where it is shown that if ϕ is invariant, the relation between $P(A_i:R_k)$ and $P(E_1:R_k)$ is convex and does not cross the main diagonal. The plots in Fig. 5, 6, and 7 clearly do not support this prediction. Indeed, estimates of ϕ for adjacent categories sometimes differed by a factor of 10. There would seem to be little hope of accounting for these data with a simple, one-parameter learning process.

Parameter estimates for the independent operator model are listed in Table 3. These provide few tests of invariance because many of the response frequencies were too extreme to give reliable values. The estimates listed give a mixed answer to the invariance question. To get a clearer evaluation of this model, the extreme values of $P(A_i:R_k)$ were set arbitrarily to .99 and .01 in the condition with $P(E_1) = .40$, and the four parameter estimates from this condition were used to generate theoretical response probabilities in the conditions with $P(E_1) = .25$ and .70. These predicted values are shown as crosses in Fig. 5 and 7. The vertical positions of the crosses, relative to the triangular data values, indicate an acceptable fit in most cases. (The horizontal positions are not predictable from the learning model and so were set to correspond to the open circles, which represent Stage C ratings.) Better estimates might well improve the agreement with data, but we show next that the random walk

TABLE 3

Parameter Estimates for the Independent Operator Model

$P(E_1)$	.25	.40	.70	.25	.40	.70
k		S_1			S_2	
1	*	.03	*	.13	.02	*
2	.16	.54	*	1.86	.43	*
3	1.47	4.10	.14	*	15.33	1.13
4	*	66.00	2.16	*	66.00	9.22
k		S_3			S_4	
1	.26	.05	*	.04	.01	*
2	1.56	.86	*	.11	.68	*
3	*	66.00	.60	*	66.00	4.67
4	*	66.00	10.35	*	66.00	14.33

Note.—Asterisks indicate estimates omitted for response probabilities less than .10 or greater than .90. The extreme values listed in the condition with $P(E_1) = .40$ were set arbitrarily for response probabilities of .01 and .99.

[4] The shape and position of these plots also conforms, qualitatively, to choice functions derived from temporal forced-choice detection experiments (W. Larkin, unpublished data). In some of these experiments, the functions were generated from "blank" trials interleaved among Signal trials, by varying the probability that A_1 would be correct. The correspondence between these plots and the data of Fig. 5, 6, and 7 may be evidence that much the same choice mechanism is involved in different psychophysical tasks.

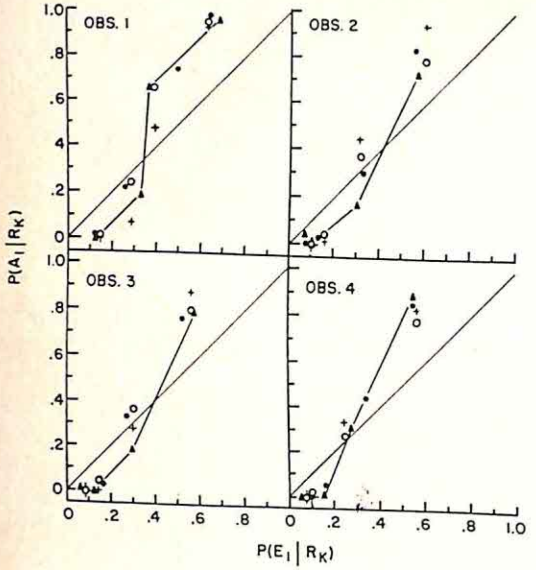

FIG. 5. Comparison of decision data with predictions from random walk and linear stochastic models for $P(E_1) = .25$ condition. (Triangular points are data from dual-response experiments. Solid dots and open circles are predictions from the random walk model, based upon Stage A and Stage C ratings, respectively. Crosses represent Equation A10. Diagonal line represents the probability-matching hypothesis.)

fit the data (triangles) as well as the learning model (crosses). However, the random walk predictions required no parameter estimates from decision data, while the learning model required four.

Sequential analyses.—The ability to account for average data does not always lead to a clear choice between theoretical formulations, as was just demonstrated for two fundamentally different conceptions of the decision mechanism. To find sharper distinctions between the learning theory and the variable criterion theory, some aspects of their sequential behavior were compared with data. One aspect of any learning model with the kind of independence assumptions used in the two just discussed is its inability to account for directional sensory generalization. Evidence of this generalization effect in the present data can be seen in Table 4, which lists estimates of $P_{n+1}(A_1:R_k)$ computed separately for each rating, R_j, given on Trial n. Because the individual sample sizes were small, data

formulation provides equivalent accuracy, with greater predictive economy.

If the detection decision depends on a criterion that undergoes a random walk, the pattern of $P(A_i:R_k)$ data can be predicted entirely from loudness ratings. Identifying the loudness categories with hypothetical sensory states, the five criterion positions are the boundaries between adjacent categories, including the extreme upper and lower positions. The distribution of these positions was estimated separately from loudness judgments at Stage A and at Stage C, and the corresponding values of $P(A_1:R_k)$ were derived for each $P(E_1)$ condition, according to equations given in the Appendix. The results are plotted in Fig. 5, 6, and 7 as solid dots (Stage A) and open circles (Stage C). The variation among these two sets of predictions can be judged in both dimensions of the plots, but the horizontal differences simply reflect the rating instability previously analyzed. In the vertical dimension, the variation is quite small, and both sets of predictions

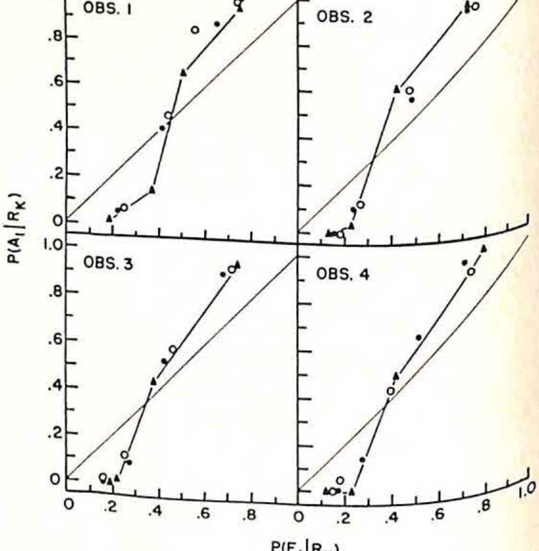

FIG. 6. Comparison of decision data with predictions from random walk and linear stochastic models for $P(E_1) = .40$ condition. (Triangular points are data from dual-response experiments. Solid dots and open circles are predictions from the random walk model, based upon Stage A and Stage C ratings, respectively. Diagonal line represents the probability-matching hypothesis. These data were used to estimate parameters for Equation A10.)

from all four *S*s have been combined. While it is risky to combine data when *S*s' choice probabilities differ, as they do here, inspection of the individual response patterns confirms that the overall results are not artifactual. The monotonicity of $P_{n+1}(A_1:R_k)$ with j is consistent with the directional generalization hypothesis. The strength of this effect seems considerable: in some rows of the table, there is a two-fold increase in the probability of A_1. Although this relation may well stem from a bias learning process, it is apparent that the independence conditions governing the two learning models are wrong for these data. A generalization scheme, such as the one described by Schoeffler (1965), seems necessary.

In its basic form, the random walk model can account for directional generalization, but it has other dynamic features that are quite vulnerable to sequential data. One of these is the assumption that response errors invariably lead to criterion shifts. A mechanism that resets the criterion less often would seem more plausible, but this refinement need not disturb the model's average properties, for, as shown in the Appendix, only the ratio of the shift probabilities affects the criterion distribution. Predictions about $P(A_i:R_k)$ are, therefore, insulated from assumptions about the overall frequency of criterion changes.

TABLE 4

ESTIMATES OF $P_{n+1}(A_1:R_k)$ CONDITIONAL ON R_j FOR TRIAL n

$P(E_1)$	k	$j = 1$	$j = 2$	$j = 3$	$j = 4$
.25	1	76	76	83	89
	2	16	25	28	40
.40	1	96	95	98	98
	2	40	52	59	64
	3	00	00	05	10
.70	3	52	64	64	72
	4	28	31	34	45

Note.—Composite data for four *S*s. Missing k values represent choice probabilities too extreme to show variation with j. Decimals are omitted from table entires. The standard error of estimate is less than .05 for nearly all entries.

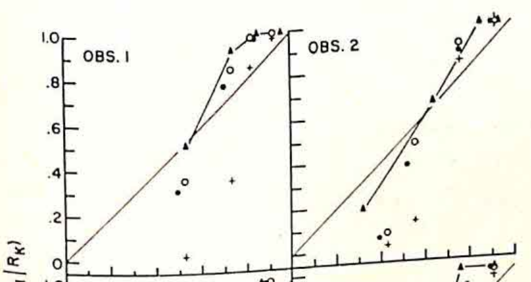

FIG. 7. Comparison of decision data with predictions from random walk and linear stochastic models, for $P(E_1) = .70$ condition. (Triangular points are data from dual-response experiments. Solid dots and open circles are predictions from the random walk model, based upon Stage A and Stage C ratings, respectively. Crosses represent Equation A10. Diagonal line represents the probability-matching hypothesis.)

To determine how well the basic model describes criterion shifts, pairs of successive trials with identical loudness ratings were isolated from the response protocols, and the A_i decision on Trial $n+1$ was examined for each combination of response and outcome on Trial n. If the criterion is stationary except after an error, these data should exhibit values of $P_{n+1}(A_1)$ that are zero or one following correct responses on Trial n. Table 5 shows the results. Again, it was necessary to combine *S*s' data, but the group pattern is representative and shows a systematic deviation from the simple criterion assumption: depending on the Signal probability, the A_i response changes on 5% to 30% of the trials that follow correct decisions. While there are several possible reasons for this mismatch with theory, two seem likely: there may be residual A_i indeterminacy after the criterion is set (equivalently, the criterion may have a random component as in Thurstonian models); or S may use a finer scale of sensory magnitudes than the four-category scale on which the analysis

TABLE 5

ESTIMATES OF $P_{n+1}(A_1)$ CONDITIONAL ON THE RESPONSE-OUTCOME EVENTS OF TRIAL n, FOR PAIRS OF SUCCESSIVE TRIALS WITH IDENTICAL LOUDNESS RATINGS

Trial n events	$P(E_1)$		
	.25	.40	.70
A_1E_1	672	839	921
A_1E_0	600	696	889
A_0E_1	109	181	339
A_0E_0	045	111	289

Note.—Composite data for four Ss. Decimals are omitted from table entries.

is based. Both possibilities suggest ways to refine the model, but they will not be pursued here.

A second important feature of Table 5 is the extent to which responses fail to change after errors. While these changes are uniformly more frequent than changes after correct decisions, the data reveal a perseveration tendancy not anticipated in the basic model. Calculations on the individual data indicate that the response criterion may shift after only 30% to 50% of the errors. (This is a range of values for the parameters α and β, defined in the Appendix, for the case $\alpha = \beta$.) When the random walk is provided with this much inertia, its behavior at the level of the A_i decisions may contradict the error-correction principle: the model may generate a higher probability for repeating an incorrect response than for repeating a correct one. The extent of this "counter-reinforcement" tendency in the individual data can be assessed in Table 6, which lists estimates of $P_{n+1}(A_1)$ conditional on the (E_i, A_j) events of Trial n. While there are large individual differences, every S shows the tendency to some extent, and all show it more strongly in the condition with $P(E_1) = .40$. This finding seems totally inconsistent with any conventional learning hypothesis, but is apparently a natural consequence of a criterion mechanism that adjusts slowly to error feedback.

TABLE 6

ESTIMATES OF $P_{n+1}(A_1)$ CONDITIONAL ON THE RESPONSE-OUTCOME EVENTS OF TRIAL n

S	Trial n events	$P(E_1)$		
		.25	.40	.70
1	A_1E_1	194	284	778
	A_1E_0	136	379	819
	A_0E_1	141	216	815
	A_0E_0	213	546	602
2	A_1E_1	204	345	754
	A_1E_0	148	434	779
	A_0E_1	155	426	769
	A_0E_0	220	487	786
3	A_1E_1	203	369	721
	A_1E_0	200	324	747
	A_0E_1	216	265	738
	A_0E_0	222	382	701
4	A_1E_1	182	370	721
	A_1E_0	250	485	688
	A_0E_1	200	268	723
	A_0E_0	242	591	691

Note.—Decimals are omitted from table entries. The standard error of estimate is less than .05 for nearly all entries.

In conclusion, we find that the empirical decomposition of a detection response into sensory and decision stages is possible to a degree that permits analysis of the way decisions are biased by experimental parameters. Overall, this finding supports the major theoretical position common to most current models (as reviewed, e.g., by Green & Swets, 1967; or by Luce, 1963a), but the detailed analyses offer little support for traditional formulations of the decision process. The likelihood-ratio hypothesis, the probability-matching hypothesis, single-state bias models, and multiple-state bias models lacking directional sensory generalization each fail to capture some important aspect of the data. There is some further suggestion that traditional learning mechanisms may be inadequate, at least if "reinforcement" is to be indentified with response feedback. The data are more efficiently, but somewhat imperfectly, described by a variable criterion theory, formalized as a random walk. While the random walk has a flexibility not always desirable in a theoretical model, the simple response mechanism it reflects may well repay further scrutiny.

REFERENCES

Atkinson, R. C. A variable sensitivity theory of signal detection. *Psychological Review*, 1963, **70**, 91–106

Atkinson, R. C., & Kinchla, R. A. A learning model for forced-choice detection experiments. *British Journal of mathematical and statistical Psychology*, 1965, **18**, 183–206.

Bekesy, G. V. *Experiments in hearing*. New York: McGraw-Hill, 1960.

Bell, D. W., & Nixon, J. C. Reliability of ratings in an auditory signal-detection experiment. *Journal of the Acoustical Society of America*, 1971, **49**, 435–439.

Blackwell, H. R. Psychophysical thresholds: Experimental studies of methods of measurement. *Bulletin of the Engineering Research Institute* University of Michigan, 1953, No. 36.

Boumann, M. A. The absolute threshold conditions for visual perception. *Journal of the Optical Society of America*, 1955, **45**, 36–43.

Bush, R. R., Luce, R. D., & Rose, R. M. Learning models for psychophysics. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford, Calif.: Stanford University Press, 1964.

Bush, R. R., & Mosteller, F. *Stochastic models for learning*. New York: Wiley, 1955.

Carterette, E. C., Friedman, M. P., & Wyman, M. J. Feedback and psychophysical variables in signal detection. *Journal of the Acoustical Society of America*, 1966, **39**, 1051–1055.

Clarke, F. R. Confidence ratings, second-choice responses, and confusion matrices in intelligibility tests. *Journal of the Acoustical Society of America*, 1960, **32**, 35–46.

Feller, W. *An introduction to probability theory and its applications*. New York: Wiley, 1957.

Garner, W. R. Context effects and the validity of loudness scales. *Journal of Experimental Psychology*, 1954, **48**, 218–224.

Green, D. M. Psychoacoustics and detection theory. *Journal of the Acoustical Society of America*, 1960, **32**, 1189–1203.

Green, D. M. Consistency of auditory detection judgments. *Psychological Review*, 1964, **71**, 392–407.

Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.

Hammond, K. R. (Ed.) *The psychology of Egon Brunswik*. New York: Holt, Rinehart & Winston, 1966.

Helson, H. *Adaptation level theory*. New York: Harper & Row, 1964.

Kac, M. A note on learning signal detection, *IRE Transactions on Information Theory*, 1962, **8**, 126–128.

Kinchla, R. A., Townsend, J., Yellott, J. I., & Atkinson, R. C. Influence of correlated visual cues on auditory signal detection. *Perception & Psychophysics*, 1966, **1**, 67–73.

Krantz, D. H. Threshold theories of signal detection. *Psychological Review*, 1969, **76**, 308–324.

Lee, W. Choosing among confusably distributed stimuli with specified likelihood ratios. *Perceptual and Motor Skills*, 1963, **16**, 445–467.

Luce, R. D. Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), Handbook of mathematical psychology. Vol. 1. New York: Wiley, 1963. (a)

Luce, R. D. A threshold theory for simple detection experiments. *Psychological Review*, 1963, **70**, 61–79. (b)

Luce, R. D., & Suppes, P. Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. 3. New York: Wiley, 1965.

McGill, W. J. Serial effects in auditory threshold judgments. *Journal of Experimental Psychology*, 1957, **53**, 297–303.

Nachmias, J., & Kocher, E. C. Visual detection and discrimination of luminance increments. *Journal of the Optical Society of America*, 1970, **60**, 382–389.

Parducci, A. Range-frequency compromise in judgment. *Psychological Monographs*, 1963, **77** (2, Whole No. 565).

Parzen, E. *Stochastic processes*. San Francisco: Holden-Day, 1962.

Schoeffler, M. S. Theory for psychophysical learning. *Journal of the Acoustical Society of America*, 1965, **37**, 1124–1133.

Shipley, E. F. Dependence of successive judgments in detection tasks: Correctness of the response. *Journal of the Acoustical Society of America*, 1961, **33**, 1142–1143.

Schulman, A. I., & Greenberg, G. Z. Operating characteristics and a priori probability of the signal. *Perception & Psychophysics*, 1970, **8**, 317–320.

Swets, J. A. Is there a sensory threshold? *Science*, 1961, **134**, 168–177.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. Decision processes in perception. *Psychological Review*, 1961, **68**, 301–340.

Tanner, W. P., Jr., & Swets, J. A. A decision-making theory of visual detection. *Psychological Review*, 1954, **61**, 401–409.

Taylor, H. A. The observing response in a cued detection task. (Tech. Rep. No. 105) Stanford, Calif.: Stanford University, Institute for Mathematical Studies in the Social Sciences, 1966.

Wickelgren, W. A. Unidimensional strength theory and component analysis of noise in absolute and comparative judgments. *Journal of Mathematical Psychology*, 1968, **5**, 102–122.

(Received June 18, 1971)

APPENDIX

A MODEL FOR CRITERION VARIATION

If detection decisions are determined by a variable criterion on an ordered set of sensory effects, the criterion changes may follow a Markov rule. The process described here is one of the simplest of this kind. A discrete time scale is provided by the experimental trial series, 1, 2, . . . , n, which is taken to be indefinitely long. A "state" of the Markov process is a criterion position; a boundary, C_k, between sensory states D_k and D_{k+1}. The S may use only a small subset of the possible sensory states if the range of stimulation is small, as it is in detection and discrimination experiments. If the range of stimulation is large, as in some absolute judgment experiments, the states D_k may be partitions or adjacent subsets of sensory effects that S treats as equivalent. In either case, the detection decision, A_i, is determined by the momentary position of the criterion: if D_j occurs on Trial n, and the criterion is C_k, then the response will be A_1 if j exceeds k, and A_0 if not. For the experiments described in this paper, sensory states are identified with loudness ratings.

A natural, though restrictive assumption is that a criterion shift may occur only after S makes an error and is informed of it by information feedback. The basic model used in this paper presumes that a criterion shift *always* occurs after an error, unless the criterion is already at an extreme position. This simplification is justified only by the relative success with which the basic model accounted for the overall results on dual-response trials. One refinement of present interest is to allow shifts from C_k to C_{k+1} with probability α, and to C_{k-1} with probability β. The parameters α and β may depend on motivational factors or they may reflect S's experience with the task: the shifts may be more frequent earlier in the experiment than later.

A second simplification restricts the criterion shifts to single steps, from k to $k+1$ or to $k-1$. Let $r_{k,n}$ denote the conditional probability that if the criterion is at Position k on Trial n, it remains there on Trial $n+1$. Similarly, let $p_{k,n}$ and $q_{k,n}$ be the probabilities that it moves to $k+1$ and $k-1$, respectively. Then $p_{k,n}$, $q_{k,n}$, and $r_{k,n}$ sum to unity. These shift probabilities depend on three factors: the "stability" parameters, α and β, the Signal probability, $P(E_1)$, and the distribution of sensory states, $P(D_k:E_i)$, for $i = 0, 1$. If four sensory states are assumed, the transition relations are:

$$\begin{aligned}
p_0 &= \alpha P(E_0) \\
p_j &= \alpha P(E_0) \cdot \sum_{k>j} P(D_k:E_0), \quad 0 < j < 4 \\
q_j &= \beta P(E_1) \cdot \sum_{k\leq j} P(D_k:E_1), \quad 0 < j < 4 \qquad \text{[A1]} \\
q_4 &= \beta P(E_1),
\end{aligned}$$

with boundary conditions $p_4 = 0$ and $q_0 = 0$. (Trial subscripts are omitted.)

These equations describe an aperiodic Markov chain with irreducible, nontransient states. Provided the sensory distributions are stationary, a unique limiting distribution of criterion probabilities, $v_k = P(C_k)$, exists (Feller, 1957, p. 356). The v_k density can be determined up to a multiplicative constant as a solution of the difference equation (see Parzen, 1962, p. 249):

$$u_j = u_{j-1}p_{j-1} + u_j r_j + u_{j+1}q_{j+1}, \quad j = 1, \cdots 3. \qquad \text{[A2]}$$

With two reflecting barriers as boundary conditions, the difference equation leads to a recursion relation:

$$u_j = u_0 \frac{p_0 p_1 \cdots p_{j-1}}{q_1 q_2 \cdots q_j}, \qquad \text{[A3]}$$

where u_0 is the undetermined constant. Equation A3 is true for any random walk on a finite state space with reflecting barriers, and, provided u_j converges absolutely, also for random walks on an infinite state space. Substituting the definitions from Equation A1, solving for u_j, and setting

$$v_j = \frac{u_j}{\sum u_i}, \quad i = 0, \ldots, 4, \qquad \text{[A4]}$$

we get the distribution of criterion positions. For computational purposes, the recursion relations in u_j may be used directly. For the present model, with u_0 arbitrarily chosen, they are:

$$\begin{aligned}
u_0 &= 1 \\
u_1 &= b/P(D_1:E_1) \\
u_2 &= u_1 b[1 - P(D_1:E_0)]/ \\
&\qquad [P(D_1:E_1) + P(D_2:E_1)] \\
u_3 &= u_2 b[P(D_3:E_0) + P(D_4:E_0)]/ \\
&\qquad [1 - P(D_4:E_1)] \qquad \text{[A5]} \\
u_4 &= u_3 b \cdot P(D_4:E_0)
\end{aligned}$$

where $b = (\alpha/\beta) \cdot P(E_0)/P(E_1)$.

If the sensory states are identified with loudness rating categories, Equation A5 provides a

way to estimate the criterion distribution from rating data. (Throughout this paper, the rating categories, R_k, are numbered in reverse order with respect to the hypothetical sensory states, D_k. Thus, R_1 should be identified with D_4; R_2 with D_3, etc.) If the ratio α/β is constant, as was assumed for the analyses in this paper, one set of rating data is sufficient to determine the criterion distribution for every Signal probability condition.

The decision probabilities, $P(A_1:D_k)$, are easily derived from the response rule. Expressed in the notation for loudness ratings, they are:

$$P(A_1:R_k) = \sum_{j=0}^{4-k} v_j. \qquad [A6]$$

Equation A6 was used to generate the predictions in Fig. 5, 6, and 7, using Stage C rating data to estimate the v_j's.

A more general version of this model has been explored by Friedman (1969), who showed that, in the Yes/No detection task, it is able to account for several of the deficiencies of fixed-criterion theories. Other suggestions for criterion variation have been made by Shipley (1961), Schoeffler (1965), Wickelgren (1968), and Nachmias and Kocher (1970).

Two learning models.—The asymptotic behavior of two learning models was used to analyze the dual-response data. The assumptions behind these models are well-known (Bush & Mosteller, 1955) and so will not be repeated here. The *independent transitions* model allows changes in $P(A_1:D_k)$ only when S is in sensory state D_k. Accordingly, the transition rule is:

$$P_{n+1}(A_1:D_k) = \begin{cases} P_n(A_1:D_k)\cdot(1-\theta)+\theta, & \text{if } D_{k,n} \text{ and } E_{1,n} \\ P_n(A_1:D_k)\cdot(1-\theta'), & \text{if } D_{k,n} \text{ and } E_{0,n} \\ P_n(A_1:D_k), & \text{otherwise.} \end{cases} \qquad [A7]$$

Here, θ and θ' are (positive) learning rate parameters, and the subscript, n, is the trial number. Essentially the same learning process was used by Kinchla, Townsend, Yellott, and Atkinson (1966) to describe the effects of cueing stimuli on detection performance. The most questionable aspect of the present application may be that a single parameter controls the learning asymptote in every sensory state. This can be shown by taking expectations on both sides of Equation A7, and letting n approach infinity. For each k, the asymptotic expected value of $P(A_1:D_k)$ is

$$P_\infty(A_1:D_k) = \frac{1}{1+\phi\dfrac{P(D_k:E_0)}{P(D_k:E_1)}}, \qquad [A8]$$

where $\phi = \theta\cdot P(E_0)/\theta P(E_1)$. Solving for ϕ, and making the appropriate substitutions, we find that

$$\frac{P(E_1:R_k)}{P(E_0:R_k)}\cdot\frac{P(A_0:R_k)}{P(A_1:R_k)}$$

is predicted to be invariant across rating categories, for any fixed Signal probability. For this to be satisfied, the relation between $P(A_1:R_k)$ and $P(E_1:R_k)$ must be a convex curve that does not cross the diagonal. Equation A8 was used to estimate ϕ from dual-response data.

The *independent operator* model permits simultaneous learning in all sensory states, but at different rates. The transition equations are:

$$P_{n+1}(A_1:D_k) = \begin{cases} P_n(A_1:D_k)\cdot(1-\theta_k)+\theta_k, & \text{if } E_{1,n} \\ P_n(A_1:D_k)\cdot(1-\theta'_k)+\theta'_k, & \text{if } E_{0,n} \end{cases} \qquad [A9]$$

Here, θ_k and θ'_k are (positive) learning rate parameters for sensory state k. Again taking expectations, and allowing n to approach infinity for each fixed k, we find the asymptotic expected value of $P(A_1:D_k)$ to be:

$$P_\infty(A_1:D_k) = \frac{1}{1+\phi_k\dfrac{P(E_0)}{P(E_1)}}, \qquad [A10]$$

where $\phi_k = \theta'_k/\theta_k$. Because ϕ_k is independent of $P(E_1)$, its invariance across Signal probability conditions provides a test of the model. The entries in Table 3 were determined from dual-response data, via Equation A10.

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 154–158

PROBE DIGIT RECALL OF ITEMS FROM TEMPORALLY ORGANIZED MEMORY SETS[1]

DONALD V. DeROSA[2] AND ROGER BAUMGARTE

Bowling Green State University

A reaction time (RT) experiment was conducted to examine the influence, on retrieval time, of temporal organization of subspan sets of digits. The *Ss* were shown five, six, or seven to-be-remembered digits followed by a probe stimulus to which they were to give the name of the item that had followed the probe in the series. Brief temporal pauses interpolated near the middle of the six- and seven-digit strings produced marked changes in the serial position curves. The results were discussed in terms of a directed memory search which was influenced by the organization of information in memory.

Recently, a number of investigators have focused attention on the influence, on retention, of the organization of information in memory. In the area of short-term memory (STM), Neisser (1967) has speculated that one form of organization is that of temporal grouping. He has suggested that *S*s actively group and partition a homogeneous string of elements and that this activity enhances the opportunity for rehearsal until a response is required. The pattern of grouped items allows the individual to assign locations to particular items in the list.

Bower and Winzenz (1969) provide evidence that the temporal organization of an auditorially presented string of 12 digits has a marked effect of retention. One of their findings indicated that transitional error probabilities (the probability that item $n + 1$ is incorrect given that item n was correct) were much higher across temporal groupings than within a grouping. It appeared that *S*s had formed units in memory which serve to segment the string into manageable subsets.

Recently, Wilkes and Kennedy (1970) demonstrated that the time to locate an item in a temporally segmented string is influenced by the pattern of grouped elements. In their study, *S*s heard a single patterned series of nine letters and then rehearsed the string aloud using the same groupings. When a criterion of three correct reproductions was reached, *S*s were shown 18 probe stimuli to which they were to report the item that had followed the probe in the list. Reactions were longest when the probe was identical to the last item of a subset.

The present investigation was conducted in order to examine the influence of temporal patterning on a visually presented subspan string of digits. The study employed the probe-recall technique used by Sternberg (1967), in which a trial consists of a serial presentation of a set of to-be-remembered items followed by a test or probe item. When a probe is shown, *S* is required to give the name of the item that followed the probe in the memorized set. Reaction time (RT) serves as the principal dependent measure. Sternberg observed behavior in this situation and concluded that *S*s used a serial self-terminating search strategy. This theory suggests that the probe is compared successively to memory representations of the items until a match is made. Then *S* retrieves the item that followed the probe and reports its name as rapidly as possible. The majority of his *S*s tended to show a linear increase in RT as a function of the serial position of the probe within the positive set. The remainder of the *S*s produced a flat curve. The former result was attributed to a serial self-terminating search which had as its starting point the

[1] This investigation was supported by National Science Foundation Grant GB-16729 awarded to Bowling Green State University and a Faculty Leaves and Research Grant from Bowling Green State University.

[2] Requests for reprints should to sent to Donald V. DeRosa, Department of Psychology, Bowling Green State University, Bowling Green, Ohio 43402.

first item of the set, while the latter result was suggestive of a search which had a random starting point. In the present investigation, special interest focused on the influence of temporal patterning on the serial position curve. If *S*s use the grouping pattern, their search strategy may be altered and this should be reflected in the serial position curves.

Method

Subjects.—The *S*s were 32 male undergraduates enrolled in introductory psychology. A first group (Group 1) of 16 *S*s served in all experimental conditions and a control condition (within-*S* control). Group 2, composed of 16 *S*s, served only as a between-*S* control and never experienced the experimental conditions.

Apparatus.—The *S*s were tested individually in a sound-deadened isolation chamber. Stimuli were digits, 1 in. in height, presented on the viewing face of a rear-projection display cell (I.E.E. Series 10) located approximately 3 ft. in front of *S*. A microphone monitored *S*s' vocal responses and also served to activate a voice-operated relay. The relay turned off the stimulus and advanced a program paper-tape reader. Reaction times were recorded by a paper-tape punch to the nearest .001 sec. A Hunter Klockounter provided a visual display of the RT.

Experimental conditions and procedure.—A memory set always consisted of five, six, or seven different digits selected from the set of single digits 1–8. Digits were white on a black background and were presented serially at a rate of 2/sec. Each stimulus was on for 200 msec. and off for 300 msec. Two seconds after the last item, a probe or test item (also a digit) was shown. This digit was distinguished as a test item by its presentation on a green background. When the probe appeared, *S* was instructed to recall the item that had followed it in the immediately preceding list. For example, if the series was "3 7 4 2 1 8" and the probe was "4," the correct response would be "2." Three seconds after *S*'s response, a new memory set was shown.

Group 1 experienced sets which, on half of the trials, had a temporal pause of 800 msec. interpolated between two digits (typically the delay between two successive items was 300 msec.). When a set contained five digits, one-half of the trials consisted of nonpartitioned ($\overline{P}$) sets and the remaining strings were divided equally between partitioned (P) strings utilizing a 2–3 grouping and a 3–2 grouping. When six items were shown, one-half of the trials consisted of a 3-3 grouping and the remaining trials were $\overline{P}$. With a set of seven items, one-half were $\overline{P}$ trials, with the remaining trials equally divided among P groupings of 3-4 and 4-3. Probes were identical to items in each serial position in the string with equal frequency. In addition, the selection of digits and their assignment to positions in the series was determined randomly. A total of 110 trials were constructed and divided equally among two lists. Each list was presented twice in two 1-hr. test sessions on 2 consecutive days.

The between-*S* control group (Group 2) was shown trials indentical to those presented to Group 1, but all trials were $\overline{P}$; that is, this group never experienced temporally organized strings. Group 2 was included as a control in order to examine the influence of experience with partitioned strings, since the control trials ($\overline{P}$) in Group 1 were experienced within the context of partitioned (P) strings.

Results

The principal analyses of the data were based on mean RTs for correct responses. Four *S*s (two in each group) evidenced unusually high error rates (all greater than 30%), and their data were not included in the analyses that follow. All other *S*s had error rates below 25% with an average of 10.6% across *S*s.

Figure 1 presents data for Group 1. Reaction time is plotted as a function of the serial position of the probe for Set Sizes 5, 6, and 7. Set size was a significant variable, $F(2, 26) = 30.69$, $p < .01$. RT increased an average of 93 msec. per unit increase in set size. This finding is consistent with previous results (Sternberg, 1967). The data for each set size were examined independently. The left panel of Fig. 1 depicts data for a set of five elements. Three curves are shown, each corresponding to one of the conditions of partitioning. The P (3–2) condition results in the fastest reactions while P (2–3) produces the slowest reactions. The $\overline{P}$ curve falls between the two P curves. Type of grouping was a reliable variable, $F(2, 26) = 5.00$, $p < .05$. Serial position of the probe was also significant, $F(3, 39) = 4.47$, $p < .01$. In addition, only the quadratic component of the serial position variable was significant, $F(1, 13) = 18.88$, $p < .01$. The Type of Grouping × Serial Position interaction was not significant, $F(6, 78) = 1.054$.

The data for Set Size 6 are shown in the middle panel of Fig. 1. The two curves represent means for the two conditions of patterning. Reactions for the P (3–3) condition were faster than reactions to $\overline{P}$ sets,

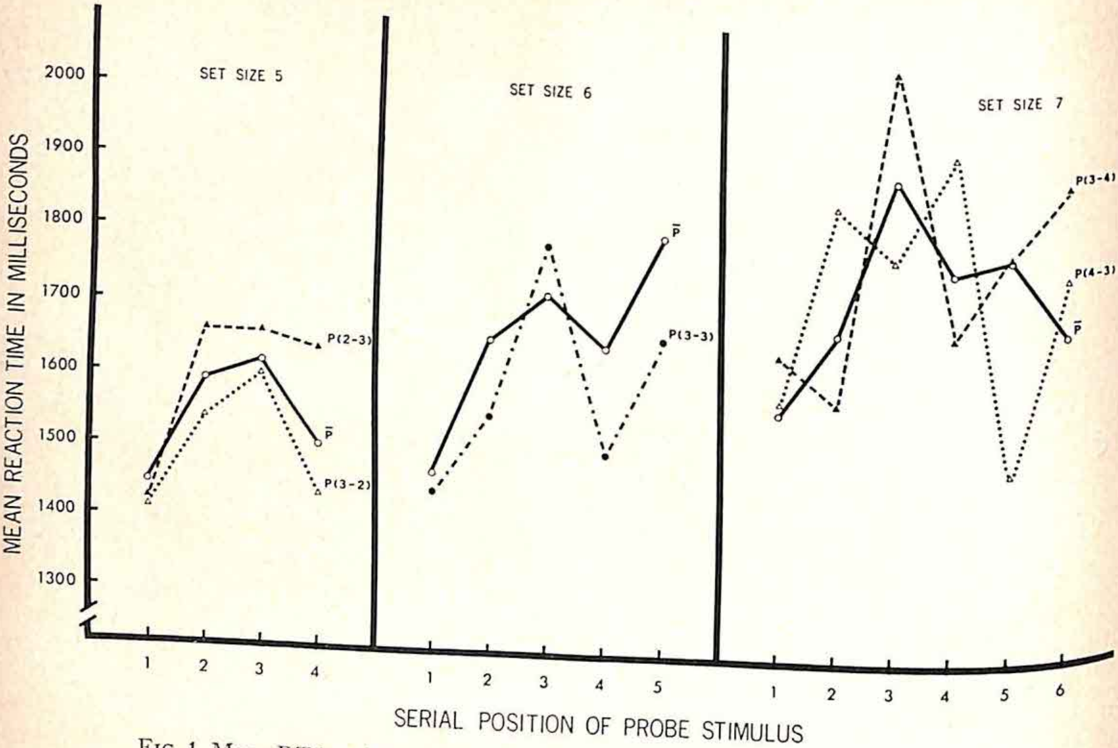

FIG. 1. Mean RT in milliseconds as a function of the serial position of the probe stimulus in the memory set for each set size.

F (1, 13) = 17.80, $p < .01$. Serial position of the probe was also a reliable variable, F (4, 52) = 7.98, $p < .01$. The two conditions produce serial position curves which differ markedly in shape. This is evident statistically in a significant Grouping Condition × Serial Position interaction, F (4, 52) = 2.62, $p < .05$. A separate analysis of the two curves indicated that serial position varied consistently under the two conditions, $p < .01$, for both. However, the trends are markedly different. For the $\overline{\text{P}}$ curve only the linear component of the variance attributable to serial position was significant, F (1, 13) = 8.83, $p < .05$. All the other higher order components were not significant. The P (3–3) curve does not have a reliable linear component, F (1, 13) = 4.52, $p > .05$, but it does have reliable quadratic and cubic components, $p < .05$, for both. The marked difference in shape of the two curves lends support to the argument that S has stored two subsets of items in memory when the set is partitioned. Furthermore, in retrieving information from memory, it appears that S can initiate his search at the beginning of either subset. The fact that the longest latency is observed for the terminal member of a subset and the fastest for initial members supports this conjecture.

Post hoc tests were conducted on some of the successive serial positions in each curve in the center panel of Fig. 1. These tests lend additional support to the hypothesis that the pauses create two subsets in memory. For example, in the P (3–3) condition, a significant increase in RT was observed between Serial Positions 2 and 3, F (1, 13) = 17.06, $p < .01$, while a significant decrease was observed between Positions 3 and 4, F (1, 13) = 8.66, $p < .01$. The same analyses comparing the identical points on the $\overline{\text{P}}$ curve yielded no reliable differences, $p > .05$, for both. For Set Size 6, the pause causes an increase in RT when the probe is identical to the item that immediately precedes the pause. A marked decrease in RT is observed when the probe is the first item following the pause. In the former case the probe and correct response are in different subsets, while in the

latter case the probe is the first member of the second subset.

The data for Set Size 7 are shown in the right panel of Fig. 1. Serial position was a significant variable, $F\ (5,\ 65) = 3.78$, $p < .01$. Type of grouping was also significant, $F\ (2,\ 26) = 4.48$, $p < .05$. The Type of Grouping $\times$ Serial Position interaction was significant, $F\ (10,\ 130) = 2.98$, $p < .01$. This result appears to be primarily due to the noticeable differences in the serial position curves for the two P sets. The longest latency for any item in the string occurs when the probe is identical to the item immediately preceding the pause. Reactions to probes identical to the item following the pause were approximately 400 msec. faster. This is evident in the marked decrease in the P (3–4) curve from Serial Position 3 to 4 and in the P (4–3) curve from Position 4 to 5.

Comparisons of the means for successive serial positions were conducted on the data for Set Size 7. In Set P (3–4) a reliable decline was observed between Positions 3 and 4, $F\ (1,\ 13) = 12.30$, $p < .01$, while the comparable comparison for the $\overline{\mathrm{P}}$ control was not significant, $F\ (1,\ 13) = 1.27$, $p > .05$. Similarly, a significant decline in RT is observed between Serial Positions 4 and 5 in the P (4–3) function, $F\ (1,\ 13) = 12.37$, $p < .01$, while the control condition does not evidence the same result, $F\ (1,\ 13) = < 1.00$.

Although the functions for the $\overline{\mathrm{P}}$ conditions in Set Size 6 and 7 do not exemplify the marked changes in slope shown by P sets, there are still noticeable departures from linearity in the $\overline{\mathrm{P}}$ control functions. For example, the $\overline{\mathrm{P}}$ curve in the right panel of Fig. 1 shows an initial increase from Position 1 to 3 and then a small but noticeable drop across the last three positions. This result, along with the small drop at Serial Position 4 for the $\overline{\mathrm{P}}$ curve in the center panel of Fig. 1, may have been due to a change in *S*'s mode of processing an unorganized set due to experience with a structured series. Winzenz and Bower (1970), for example, have shown that *S*s trained on one standard grouping format recoded different grouped strings into the original training format.

In order to examine the possibility that *S*s grouped the $\overline{\mathrm{P}}$ sets due to practice with organized sets data from the between-*S* control was compared with the data ob-

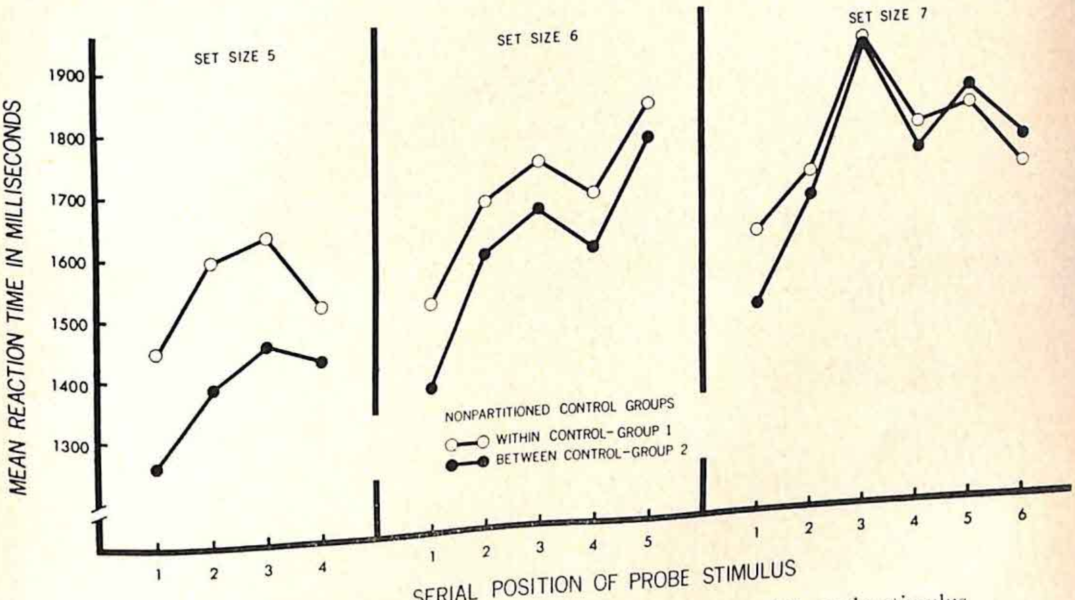

FIG. 2. Mean RT in milliseconds as a function of serial position of the probe stimulus in the memory set for within-*S* $\overline{\mathrm{P}}$ control and between-*S* $\overline{\mathrm{P}}$ control.

tained from the within-S control group shown in Fig. 1. This comparison is depicted graphically in Fig. 2. The most striking feature of the data in Fig. 2 is the marked similarity in shape of the two functions in each panel of Fig. 2. It is clear from the data obtained from the between-S control group that the departures from linearity observed for the within-S control were not the result of experience with organized strings. The Ss in the between-S control group exhibit serial position curves which are essentially identical to the comparable functions observed for the within-S control. Analyses conducted on the functions in each panel of Fig. 2 indicated the following results for all set sizes: (*a*) serial position was a reliable variable, (*b*) no significant difference between groups, and (*c*) the absence of any reliable Serial Position $\times$ Groups interaction. In only one case does a statistical test between the two control groups even approach significance. This occurs in Set Size 5 where reactions for the between-S control group are 162 msec. faster than the mean for the within-S control. The difference is not statistically significant due to the large within groups variance, $F(1, 26) = 3.22$, $p > .05$.

Discussion

The results of this investigation demonstrate that temporal grouping of a visually presented subspan set of integers influenced retrieval strategies. This is especially true for sets of six and seven items. With a small set (five items), no consistent positional differences were observed as a function of the particular grouping used. With larger sets, the utilization of the pattern becomes more apparent. This is reasonable since larger sets place a greater burden on STM and rhythmic patterning should be most helpful with these sets.

Grouping was not uniformally beneficial to all items within the memory set. RT was not generally faster at all locations in a P condition relative to $\overline{P}$ sets. The typical finding demonstrated that grouping slowed retrieval when the probe item was the terminal member of a subset. On the other hand, RTs to the first member of the second subset were enhanced. This result is consistent with a retrieval strategy which is directed and takes advantage of the organization present in the input sequence. With P sets it may have been possible for Ss to initiate a search with the first member of a subset instead of with the first item in the total set. This directed search would explain the rapid reactions to initial elements of a subset and the typical increase in RT to elements experienced later in the subset.

References

BOWER, G. H., & WINZENZ, D. Group structure, coding, and memory for digit series. *Journal of Experimental Psychology*, 1969, **80**(2, Pt. 2).

NEISSER, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

STERNBERG, S. Retrieval of contextual information from memory. *Psychonomic Science*, 1967, **8**, 55–56.

WILKES, A. L., & KENNEDY, R. A. The relative accessibility of list items within different pause-defined groups. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 197–201.

WINZENZ, D., & BOWER, G. H. Subject-imposed coding and memory for digit series. *Journal of Experimental Psychology*, 1970, **83**, 52–56.

(Received June 21, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 159–161

SHORT REPORTS

EFFECTS OF BLOCKING OF INPUT AND BLOCKING OF RETRIEVAL CUES ON FREE RECALL LEARNING

SUSAN P. LUEK,[1] JOHN P. McLAUGHLIN, AND GEORGE A. CICALA
University of Delaware

To evaluate an independent versus a dependent storage model for memory, the effects of ordering of categorized input words and retrieval cues on free recall learning were assessed. An 80-word categorized list was presented in blocked and unblocked orders for three study-test trials. Half-list cues were either blocked, unblocked, or absent at recall. The results indicated that blocked input facilitated recall. With blocked input, both blocked and unblocked cues facilitated recall. With unblocked input, blocked cues did not facilitate recall, and unblocked cues decreased recall. An interpretation in terms of the interdependence of stored items was suggested.

In a series of studies, Slamecka (1968, 1969) has provided evidence supportive of an independent storage model for memory. His characteristic procedure has been to present *Ss* with acquisition-list words as retrieval cues in free recall learning and to require the recall of critical remaining words. Slamecka found that cueing did not facilitate recall over that of control *Ss* who received no cues. In that cueing failed to facilitate recall under numerous presentation conditions, Slamecka has argued for the generality of the independent storage model.

A finding most damaging to a clustering or interdependent memory model is Slamecka's (1968) demonstration that the recall of taxonomically related materials was not facilitated by cueing. Thus, with ample opportunity for the organization of input, organization seemingly did not occur.

Results of subsequent experiments by Wood (1969), Hudson and Austin (1970), Lewis (1971), and others have challenged this conclusion. In addition to providing an opportunity for organizing input by presenting related words, Wood maximized the probability of organization by structuring lists by blocking related items. He found that cueing facilitated recall only when a blocked presentation of items was used. Lewis verified the results of Wood using taxonomically related materials. Hudson and Austin demonstrated that the facilitating effects of the cueing of taxonomically related materials was not dependent upon the blocked presentation of related items. However, the procedure of both Hudson and Austin differed from that of both Wood and Lewis in that prior to the experiment, Hudson and Austin told their *Ss* the exact number of categories in the list and the names of the categories. It thus appears that when the categorical structure of materials is made evident, either by the blocking of related items during input or by specific instructions, items are stored consonant with that structure and do interact.

The present experiment was done to assess the effects of the blocking of related items during input to establish the phenomenon demonstrated by Wood (1969) and Lewis (1971). It was further reasoned that with blocked presentation the blocking of recall cues would facilitate recall, since it might be expected that blocked cues would elicit a category search strategy which would be effective if items were stored in categories. Further, if, with unblocked presentation, items were stored by category, the blocked cues should facilitate recall when items were not blocked during presentation. Therefore, the effects of blocked or unblocked cues, when input was either blocked or unblocked, were assessed.

Method.—The *Ss* were 120 students enrolled in various sections of introductory psychology at the University of Delaware. Their participation partially fulfilled a course requirement. All *Ss* were naïve to verbal learning experiments.

Four words were chosen from each of 20 randomly selected categories of the extended Connecticut category norms of Battig and Montague (1969). Each word was one of the eight most frequently listed for each category, and the mean frequency of emission of the 80 words by the combined Maryland and Illinois samples was 270.14.

All *Ss* were given three alternating study-test trials on the list. The presentation order of the words was either blocked (B) according to intact categories or unblocked (U), such that no two words from a category occurred consecutively. Both the order of the categories and the sequence of the words within categories remained constant from trial to trial in the B presentation order. The U presentation order was also constant across trials.

The retrieval cues consisted of 40 list words, 2 words from each of the 20 categories. These cue words were either unblocked (U) in a random sequence which was in no way related to input order or blocked (B) into 2-word categories which reflected the sequence of the B input order. The B cues consisted of the third and fourth items from each input category block. (For example, if the blocked input order of the category, musical instruments, was "drum, trumpet, flute, guitar," the B cue order would be "flute, guitar"). The U and B cue orders were typed on sheets which were in-

[1] Requests for reprints should be sent to Susan P. Luek, Department of Psychology, Wolf Hall, University of Delaware, Newark, Delaware 19711.

cluded in the recall booklets of the cued experimental groups.

The four experimental groups were differentiated on the basis of (*a*) the presentation order of the list words during the study trials and (*b*) the presentation order of list-word cues at recall. These conditions were labeled BB, BU, UB, and UU, where the first letter indicates the input order and the second letter indicates the order of cues. (For example, *S*s in Group BU received the blocked presentation order during each study trial and the unblocked ordering of list-word cues at each recall). The task of the experimental groups was to recall the 40 remaining words (critical words) which did not appear as cues.

The two control (C) conditions received no cues at recall. They were called the blocked control (BC) and the unblocked control (UC), based upon the presentation order during each input. The *S*s in the control groups were instructed to recall as many of the 80 words as possible during the recall periods.

Testing was conducted at eight scheduled sessions. The *S*s were permitted to report to any one of the sessions, and the number of *S*s per session ranged from 7 to 38. The unblocked presentation order was used at four of the testing sessions, and the blocked presentation order at the other four, in an alternating sequence.

Upon arrival at the testing session, each *S* was randomly assigned to one of the three cueing conditions and the presentation order being tested at that session. This random assignment was continued at each session such that 20 *S*s were in each of the six conditions of the experiment. The conditions contained approximately equal numbers of males and females.

At the beginning of the sessions all *S*s were given typed instruction sheets and the booklets for written recall. The *E* verbally summarized the instructions just prior to the first presentation of the list. All items were presented to *S*s with a Wollensack tape recorder at a 2-sec. rate. The recall period which followed each presentation of the list was 6 min.

Results and discussion.—All groups were scored on the recall of critical words. Figure 1 shows that the groups were ordered in terms of decreasing critical word recall as follows: BB, BU, BC, UC, UB, and UU. An analysis of variance was done on the recall scores for these groups across the three test trials, and this analysis showed the groups to differ significantly, $F\ (5,\ 114) = 21.91$, $p < .001$. The expected trials effect was significant, $F\ (2,\ 228) = 775.52$, $p < .001$. The interaction did not approach significance.

All group comparisons of critical recall were made with a series of Duncan multiple-range tests. As can be seen from Fig. 1, blocked presentation of list words during input produced uniformly greater recall than unblocked input, in that all three functions for blocked presentation are higher than those for unblocked presentation. (All comparisons were significant at least at the .05 level). These results emphatically support those obtained by Wood (1969) and Lewis (1971), which showed blocked input to be superior to unblocked with cued recall. More germane to the present problem are the effects of cueing with blocked or unblocked cues when input was either blocked or unblocked. When input was blocked, recall was ordered such that blocked retrieval cues produced the greatest recall, no retrieval cues produced the least recall, and unblocked retrieval cues fell intermediate to the other two conditions (BB > BU and BC, $p < .001$; BU > BC, $p < .05$).

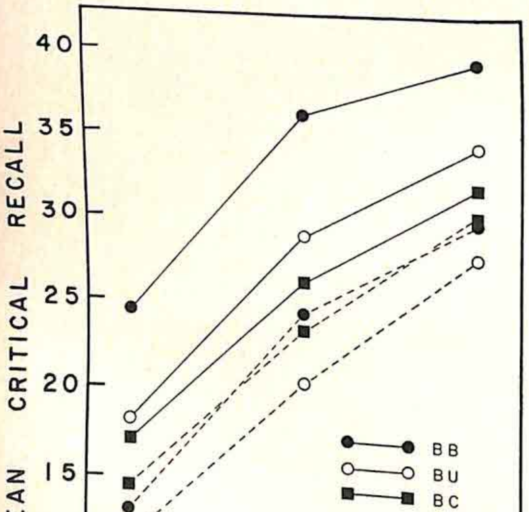

FIG. 1. Mean critical recall as a function of trials. (The solid lines represent the blocked input for each recall condition. The dotted lines represent the unblocked input groups.)

The data clearly indicate that with categorized materials in which categories are made salient by blocking during input, a memory model specifying the independent storage of items is totally inappropriate. That blocked cues greatly facilitated recall can only mean that the items in storage interacted with the cues presented during recall. Indeed, even the presentation of unblocked cues facilitated recall over that obtained with the noncued condition. This finding lends further support to the view that items in memory interact. Slamecka (1968) carefully specified an independent storage model as one which permits the prediction that the facility with which one item is recalled does not alter the accessibility of other items. The data from the blocked input conditions of the present experiment do not support this prediction.

The results of the unblocked input groups indicated that blocking of cues did not facilitate recall

over that of the noncued control group, a finding which might be interpreted as favorable to the independent storage model. If a dependent storage model is appropriate and if *S*s were storing the items in the preconceived categories, the blocked cues should have facilitated recall. It may be that merely structuring lists with E-determined categories does not guarantee that *S*s will identify these categories and use them in the storage process, particularly when so few learning trials are permitted.

A serious interpretive problem for the independent storage model is posed by the finding that with unblocked presentation, unblocked cues produced less recall than either blocked cues or no cues (UU < UC and UB, $p < .001$). If items were stored independently, there is no immediately apparent reason why randomly selected items should have interfered with recall. Random cues may have produced a decrement in recall by eliciting categories which were not consonant with the idiosyncratic organization imposed upon the materials during storage.

Taken together, the results of this experiment are interpreted as supporting a categorical structure of memory in which stored items are organized together and are interdependent.

REFERENCES

BATTIG, W. F., & MONTAGUE, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, 1969, **80** (3, Pt. 2).

HUDSON, R. L., & AUSTIN, J. B. Effect of context and category name on the recall of categorized word lists. *Journal of Experimental Psychology*, 1970, **86**, 43–47.

LEWIS, M. Q. Categorized lists and cued recall. *Journal of Experimental Psychology*, 1971, **87**, 129–131.

SLAMECKA, N. J. An examination of trace storage in free recall. *Journal of Experimental Psychology*, 1968, **76**, 504–513.

SLAMECKA, N. J. Testing for associative storage in multitrial free recall. *Journal of Experimental Psychology*, 1969, **81**, 557–560.

WOOD, G. Retrieval cues and accessibility of higher-order memory units in multitrial free recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 782–789.

(Received April 1, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 161–164

TEST OF THE PROPRIETY OF THE TRADITIONAL DISCRIMINATIVE CONTROL PROCEDURE IN PAVLOVIAN ELECTRODERMAL AND PLETHYSMOGRAPHIC CONDITIONING[1]

JOHN J. FUREDY AND KARL SCHIFFMAN[2]

University of Toronto

According to R. A. Rescorla's contingency position, the traditionally used explicitly unpaired CS− (euCS−) of the differential-conditioning paradigm is an inappropriate control for CS+ conditioning. It follows that the CS+: euCS− performance difference should exceed the difference between CS+ performance and performance to a truly random CS− (trCS−), a stimulus which is uncorrelated with UCS occurrence. This implication was tested on 32 *S*s in a .75-sec. delay conditioning study with GSR and digital volume pulse change as dependent autonomic variables. Reliable autonomic discrimination between CS+ and the two control CSs (euCS− and trCS−) was obtained, but the implication derived from the contingency position was not confirmed in either autonomic measure, even though *S*s were shown to be aware of the contingency differences considered important by Rescorla.

The contingency account of Pavlovian conditioning, as formulated by Rescorla in a methodological (1967) and an empirical review (1969) paper, makes two basic assertions. One is that the crucial factor in Pavlovian conditioning is CS-UCS contingency rather than CS-UCS pairing (Rescorla, 1967, pp. 75–76). The other and related claim is that "conditioned inhibition has a status equal to that of conditioned excitation [Rescorla, 1969, p. 93]," where inhibition and excitation are assumed to be generated by introducing, respectively, negative and positive contingencies between a CS and a UCS.

An implication of Rescorla's influential position is that in the traditional discriminative conditioning paradigm the CS− is an inappropriate control for assessing CS+ conditioning (Rescorla, 1967, p. 74). However, while Rescorla's stricture against the discrimination procedure follows most clearly and logically from the contingency position, it is a separate question whether his argument has empirical im-

[1] This research, a short version of which was reported at the meeting of the Midwestern Psychological Association, Detroit, May 1971, was supported by grants from the National Research Council (APA 222) and from Canada Council (S70-0710) to JJF. We are indebted to L. de Toledo for critical advice concerning some of the ideas in this paper, although we should add that on some aspects we have finally had to agree to disagree.

[2] Requests for reprints should be sent to John J. Furedy, Department of Psychology, University of Toronto, Toronto 181, Canada.

portance for autonomic behavior. The conditions for testing this question are obtained from a two-groups differential-conditioning design in which one group receives the conventional, explicitly unpaired CS− (euCS−), while the other group receives the Rescorla-recommended truly random CS− (trCS−), the occurrence of which is not negatively correlated, but is merely uncorrelated, with the occurrence of the UCS. Provided discrimination is obtained in the form of CS+ performance exceeding CS− performance, it follows from the contingency position that the CS+:euCS− performance difference should exceed the CS+:trCS− performance difference, and it is this autonomic implication of the contingency position that the present study sought to examine. Moreover, to cast more light on the contingency position in the context of human conditioning, advantage was taken of the verbalizing ability of the present *S*s by obtaining postexperimental measures of the subjective contingencies between the CSs and the UCS.

Method—The *S*s were 32 undergraduates from the University of Toronto, who participated either to fulfil a course requirement or for $1.50.

Details of the general apparatus are available in Furedy (1970b), while the apparatus used here was identical to that described by Furedy (1971).

The procedure employed a .75-sec. delay-conditioning paradigm with durations of .75, .75, and .3 sec. for the CS-UCS interval, CSs (tone and light), and UCS (2.0-mA shock), respectively. This arrangement is similar to one that has yielded highly reliable differential conditioning in a previous study (Furedy, 1970a, "short-forward" condition). Before conditioning, all *S*s received a series of two tone-CS and two light-CS trials interspersed with three UCS-alone trials; the intervals between these seven preliminary trials varied randomly between 30, 40, and 50 sec. As in Furedy (1970a), the conditioning trial series for the 16 *S*s in the euCS− group comprised 12 pairings of CS+ with UCS, 15 euCS− trials, and three test CS+ alone trials, presented, respectively, after 0, 6, and 12 CS + UCS trials. From *S*'s viewpoint, the nature of the trials was unsystematically ordered, and the intertrial intervals varied randomly between 30, 40, and 50 sec. For the remaining 16 *S*s in the trCS− group, the conditioning-trial series was the same as for the euCS− group, except that the 15 euCS− trials were replaced by 15 trCS− trials. These trCS− trials, as detailed in Furedy (1971), were randomly allocated over the total conditioning session. Hence, although UCS occurrence was not random through the session, as was the case in Rescorla's (1967, 1969) experiments, the occurrence of trCS− in the present study was independent of, or uncorrelated with, UCS occurrence. Specifically, in contrast to CS+, which was followed by a UCS within .75 sec. on 12 of 15 occasions, and to euCS−, which was never followed by a UCS within 29 sec., the allocation of the 15 trCS− trials in this experiment resulted in 3, 3, 0, and 9 occasions, respectively, in which a UCS followed a trCS− within 0–10, 10–20, 20–29, and 29–109+ sec., with the last trCS− occurring after the last UCS. Within each of the euCS− and trCS− groups, the nature of CS+ (light versus tone) was counterbalanced between *S*s.

In addition to the autonomic GSR and peripheral vasoconstriction (VPC), measures which were continuously monitored and recorded throughout the session, subjective contingency (SC) measures were obtained at the end of the experiment by a written questionnaire which was prefaced by a written and diagrammatically illustrated explanation of the concept of a continuum which runs from negative through zero to positive contingency. After attempting to ensure that *S* had understood this concept, *E* asked *S* to give magnitude estimates (by marking a point along a 5-in. line) of the contingencies during the conditioning stage between the tone (CS+ or CS−) and the light (CS− or CS+) on the one hand and the shock (UCS) on the other hand.

Results.—The GSR, a drop in resistance from 1 to 5 sec. following stimulus onset, was defined as the resistance difference between the point of response initiation and minimal resistance level reached during the 1–5 sec. interval. These resistance changes were transformed into micromho, conductance-change units. The VPC was defined within the 2–8 sec. latency range following CS onset, as in a previous study (Furedy, 1971), and was transformed into percent-VPC units.

As regards autonomic discrimination (CS+ performance exceeding CS− performance), the GSR preliminary-trials data yielded no significant difference between CS+ performance ($\bar{X}$ = 2.12) and CS− performance ($\bar{X}$ = 1.51), F (1, 28) = 1.94, $p > .1$. However, the conditioning-trials GSR data indicated highly reliable discrimination, with CS+ performance ($\bar{X}$ = 3.26) exceeding CS− performance ($\bar{X}$ = 2.33), F (1, 28) = 12.24, $p < .01$. In addition, there was a Discrimination × Trials interaction with a divergence over trials of the CS+ function ($\bar{X}$ = 2.14, 3.84, and 3.94, respectively, for first, second, and third CS+ test trial) and the CS− function ($\bar{X}$ = 1.94, 2.67, and 2.37, respectively, for the three CS− trials proximal to each of the three CS+ test trials), F (2, 56) = 7.52, $p < .01$. The VPC pattern of results, although markedly more variable than the GSR, also suggested that the conditioning operations had developed reliable discrimination. The preliminary-trials difference between CS+ performance ($\bar{X}$ = 24.9) and CS− performance ($\bar{X}$ = 23.5) did not approach significance, $F < 1$, while during conditioning CS+ performance ($\bar{X}$ = 30.0) significantly exceeded CS− performance ($\bar{X}$ = 23.9), F (1, 28) = 7.28. As for the GSR, there was divergence over conditioning trials of the CS+ ($\bar{X}$ = 30.0, 27.1, and 32.5, respectively, for the first, second, and third CS+ test trial) and CS− functions ($\bar{X}$ = 28.5, 21.7, and 21.4); but, in contrast to the GSR, this divergence in the VPC, as represented by the Discrimination × Trials interaction effect, was not reliable, F (2, 56) = 1.56, $p > .1$.

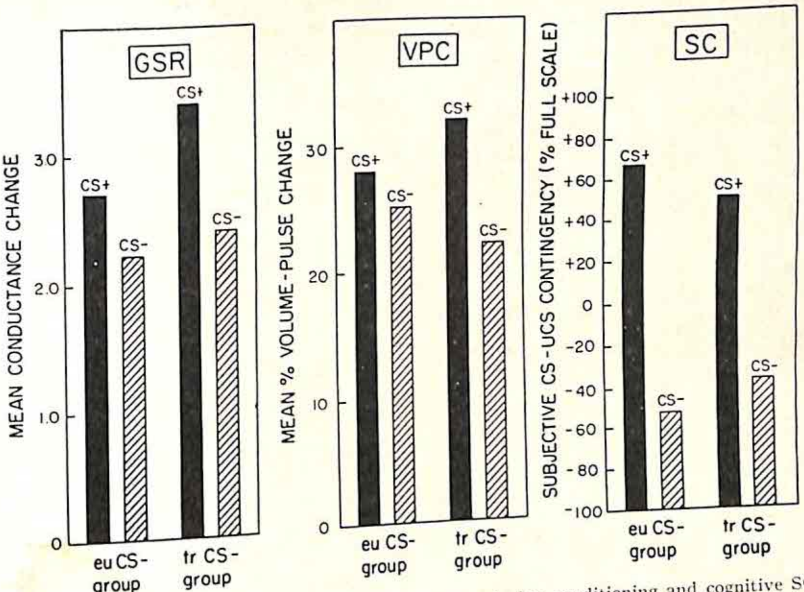

FIG. 1. Autonomic GSR (left panel) and VPC (center panel) measures during conditioning and cognitive SC (right panel) measures after conditioning on CS+ and control CS− trials in the euCS− ($N = 16$) and trCS− ($N = 16$) groups.

The left (GSR) and center (VPC) panels of Fig. 1 show the data relevant to the prediction derived from the contingency position, that CS+:CS− autonomic discrimination in the euCS− group exceed that in the trCS− group. Inspection of the panels indicates that even the trends were contrary to the contingency expectations both for GSR and VPC, since, in both measures, the mean algebraic CS+:CS− performance difference in the trCS− group exceeded that in the euCS− group. These trends, however, were not reliable for either measure, $t\ (30) = 1.48$ and 1.68 for GSR and VPC, respectively. Further inspection of the GSR and VPC panels in Fig. 1 could suggest the possibility that especially in the case of the GSR, the failure to obtain a greater CS+:CS− difference during conditioning in the euCS− group may have been due to an abnormally low CS+ performance level in that group relative to the CS+ performance level in the trCS− group. However, apart from the question of whether such an explanation would really aid the contingency position, statistical analysis did not support the explanation, since, even in the GSR, the difference between the CS+ performance levels during conditioning between the euCS− and trCS− groups did not approach significance, $t = .89$.

The SC results are shown in the right-hand panel of Fig. 1. As for the autonomic measures, clear cognitive discrimination of the CS-UCS contingencies was obtained as regards the CS+:CS− difference, with CS+ being judged as having been more positively correlated with UCS occurrence than was CS− by all but 1 of the 32 *Ss*, $t\ (31) = 11.3$, $p < .001$. However, in contrast to the autonomic trends, the CS+:CS− difference in the euCS− group did exceed the same difference in the trCS− group, as predicted by the contingency position and as expected on the basis of the actual CS−UCS contingencies in the experiment. These SC trends were statistically confirmed in two respects: (a) comparison of the CS+:CS− algebraic SC differences between the euCS− and trCS− groups yielded $t\ (30) = 1.77$, $p < .05$, one-tailed; (b) the same comparison, but without the deviant single *S* who failed to rate the contingency of CS+ as higher than CS− (probably because of having failed to understand the contingency concept), yielded $t\ (29) = 2.53$, $p < .01$, one-tailed.

To assess the relationships between the extents of autonomic and cognitive discrimination, relationships which are relevant to recent and commonly accepted claims that "verbalized knowledge of CS−UCS relations is related to the extent of differential conditioning [Baer & Fuhrer, 1970, p. 178]," product-moment correlations were computed between pairs of measures on the algebraic CS+:CS− difference scores. Neither electrodermal (GSR) nor plethysmographic (VPC) discrimination was correlated with cognitive (SC) discrimination, $r = -.00$ and $-.02$ for GSR−SC and VPC−SC, respectively. However, the two autonomic measures of discrimination, GSR and VPC, were positively and significantly correlated, $r = .41$, $p < .05$.

Discussion.—The presence of reliable discrimination in both autonomic measures allowed the prediction derived from the contingency position to be adequately tested in a situation which, unlike previous experiments (Furedy, 1971), contained no such idiosyncratic features as UCS-alone trials. Yet, as in those experiments, the present autonomic behavior did not conform to contingency expectations. Not only did the discrimination (CS+ versus CS− performance) in the euCS− group fail to significantly exceed that in the trCS− group, but even the trends in both GSR and VPC (Fig. 1)

were in the opposite direction to that predicted by the contingency position.[3]

It is possible that the apparent contradiction between the present anticontingency autonomic results and the procontingency results quoted by Rescorla (1969) is resolvable in terms of the relatively few CS+UCS training trials used in the present experiment. To say that, however, is still to deny the claim that Rescorla's stricture against the traditional discriminative control procedure applies to autonomic conditioning experiments, since these, almost without exception, have employed short acquisition periods with relatively few conditioning trials. In the present case, moreover, the acquisition period seems to have been sufficiently long to allow the human *S*s to become demonstrably sensitive in terms of verbal report to those contingency differences considered important by Rescorla. Yet, despite this sensitivity, the autonomic measures, though showing reliable differential conditioning, did not yield the pattern of results predicted by the contingency position, and, as in the previous experiments (Furedy, 1971), no evidence was obtained for the presence of any conditioned autonomic CS− inhibition (Rescorla 1967, 1969).

[3] In defence of the contingency position, it is true, as pointed out by a consultant, that the ordering of mean response magnitudes during conditioning obtained for the GSR (Fig. 1) is one which would be predicted by the contingency position together with considerations of interstimulus generalization between CS+ and CS−. At first sight this pattern of ordering (trCS+ > euCS+ > trCS− > euCS−) might be considered to be strong evidence for the contingency position, if one accepts the view that this particular order has an approximate probability of occurrence of .04. However, given the constraint of discrimination (CS+ > CS−), there are only four possible patterns of ordering, and the appropriate *p* value for the obtained GSR pattern becomes .25. Moreover, the VPC data (Fig. 1) yielded an ordering of trCS+ > euCS+ > euCS− > trCS−, while GSR data from the previous study (Furedy, 1971, Exp. I), where reliable discrimination was obtained, yielded an ordering of euCS+ > trCS+ > euCS− > trCS−. Arguments based on such orderings, therefore, when closely examined, do not appear to support the contingency position. Another argument, for which we are indebted to L. de Toledo, is that the presence of the CS+ trials before each UCS could lead *S* to distinguish between those trCS− trials which occur in compound with the CS+ (trCS− and proximal CS+) and those trCS− trials which occur apart from the CS+ (trCS− and no proximal CS+). In this case, the latter class of trCS− trials (to which autonomic responding was measured) could acquire inhibitory properties, since they would never be immediately followed by a UCS. However, even if the contingency between the present trCS− and the UCS was not truly zero for this reason, the contingency position should still predict that the trCS− would have been less negative than the euCS−, since the latter stimulus was never followed by a UCS within 29 sec., whereas the former, even in the absence of a proximal CS+ trial, was thrice followed by a UCS within 10–20 sec. Hence, the contingency position still predicts a greater CS+:CS− difference in the euCS− group than in the trCS− group, a difference which did emerge reliably in terms of subjective contingency, but not in the autonomic measures. Nevertheless, it is likely that to obtain a trCS− where the contingency between the trCS− and the UCS is clearly zero, separate groups will have to be run with CS+, trCS−, and euCS− as the CSs in the three-groups, rather than the two-groups, discrimination design that was employed here.

REFERENCES

Baer, P. E., & Fuhrer, M. J. Cognitive processes in the differential trace conditioning of electrodermal and vasomotor activity. *Journal of Experimental Psychology*, 1970, **84**, 176–178.

Furedy, J. J. CS and UCS intervals and orders in human autonomic classical differential trace conditioning. *Canadian Journal of Psychology*, 1970, **24**, 417–426. (a)

Furedy, J. J. Test of the preparatory-adaptive-response interpretation of aversive classical autonomic conditioning. *Journal of Experimental Psychology*, 1970, **84**, 301–307. (b)

Furedy, J. J. Explicitly-unpaired and truly-random CS− controls in human classical differential autonomic conditioning. *Psychophysiology*, 1971, in press.

Rescorla, R. A. Pavlovian conditioning and its proper control procedures. *Psychological Review*, 1967, **74**, 71–80.

Rescorla, R. A. Pavlovian conditioned inhibition. *Psychological Bulletin*, 1969, **72**, 77–94.

(Received April 2, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 1, 165–167

COMPARISON OF TRACE AND DELAY CLASSICAL EYELID CONDITIONING AS A FUNCTION OF INTERSTIMULUS INTERVAL[1]

SUSAN M. ROSS[2] AND LEONARD E. ROSS
University of Wisconsin

Trace and delay classical eyelid conditioning were compared in adult human *S*s at seven interstimulus intervals from 250 msec. to 1,400 msec. Pure-tone CSs were used with the trace CS having a duration of 50 msec. and the delay CS terminating with the UCS, a puff of compressed air to the right cornea. The results indicated no significant trace-delay difference in conditioning performance.

Although it is generally assumed that trace and delay CSs are equally effective in the single-cue eyelid conditioning of humans, this assumption is based upon comparisons between studies rather than upon a direct trace-delay comparison within a single study with other variables kept constant. Thus, the combined data of studies by Reynolds (1945), Kimble (1947), and McAllister (1953) are often cited as evidence for the equivalence of trace and delay CSs although there are a number of procedural differences which make the direct comparison of the results of these studies inconclusive. For example, a nonreinforced test trial procedure was used in both of the delay conditioning studies (Kimble, 1947; McAllister, 1953) but not in the trace conditioning study (Reynolds, 1945), a confounding of the use of test trials, and therefore partial reinforcement, with the trace-delay variable which might have reduced the level of performance of the delay conditioning groups relative to that of the trace groups. Other differences among the studies involved the CS employed (lights, tones, or clicks), CS duration, and the presence or absence of a ready signal preceding each trial.

In view of the above, and the fact that trace-delay differences have been found in the classical conditioning of other organisms as a function of the interstimulus interval (ISI) (Schneiderman, 1966), the present two experiments were conducted to provide a direct comparison of trace and delay procedures at several ISIs. In the first study, trace and delay conditioning were compared at four ISI values, 500 msec., 800 msec., 1,100 msec., and 1,400 msec. These ISIs were selected in order to include a value which is generally accepted as the optimal for single-cue conditioning in adults, as well as several longer intervals. In the second study, conducted 1 yr. later, trace and delay procedures were compared at ISIs of 250 msec., 300 msec., 350 msec., and 500 msec. in order to extend trace-delay comparisons to short ISIs. The first and second study overlapped at one ISI value, 500 msec., to indicate the comparability of the two *S* samples.

Method.—Eighty-eight *S*s participated in Exp. I, and 96 *S*s served in Exp. II. All *S*s were students in an introductory psychology course at the University of Wisconsin, and all earned "points" which counted toward their course grade.

The studies were conducted in a three-room laboratory which permitted simultaneous conditioning of two *S*s. All sound-generating equipment and response-recording equipment was located in one of the three rooms (see Wilcox & Ross, 1969, for a complete description of this equipment), and one *S* was seated in each of the other two rooms. Throughout the conditioning session, a silent motion picture was projected on a screen in front of *S*.

Earphones were used to present a 65-db. white-noise background and the two pure-tone CSs (800 Hz. and 2,100 Hz.) which were matched in loudness to a 1,000-Hz. tone with an intensity of 85-db. SL (see Wilcox & Ross, 1969, for exact specification of these stimuli). All tones had a "fast" rise and decay time ($\approx$10 usec.). The trace CS had a duration of 50 msec., and the delay CS terminated with UCS offset. For both conditions, the ISI was measured from the onset of the CS to the onset of the UCS. The UCS was a puff of compressed air to the right cornea with an intensity of .75 psi and a duration of 100 msec. Intertrial intervals of 15, 20, and 25 sec. with a mean of 20 sec. were employed.

Experiment I included trace-delay comparisons at ISIs of 500, 800, 1,100, and 1,400 msec. In Exp. II trace-delay comparisons were made at ISIs of 250, 300, 350, and 500 msec. In both studies, half of the *S*s in each group were conditioned with the 2,100-Hz. tone and half were conditioned with the 800-Hz. tone. In addition, half of the *S*s in each experimental group were males and half were females.

While *S* was being seated in his experimental room, he was informed that the study was concerned with his reactions to certain events as these events occurred while he was watching a silent movie. After the headband and earphones had been fitted on both *S*s, further instructions were presented through the earphones. The *S* was asked to refrain from touching the equipment on his head and to relax, sit quietly, watch the movie, and make no attempt to aid or inhibit his natural responding. Two CS-alone trials and a trial with the air puff alone were presented to each *S* with the instructions that such events would occur while he was watching the film. The film was then started, and each *S* was given 80 single-cue conditioning trials.

[1] This investigation was supported by United States Public Health Service Research Grant MH 10235.

[2] Requests for reprints should be sent to Susan M. Ross, Department of Psychology, Psychology Building, University of Wisconsin, Madison, Wisconsin 53706.

Results and discussion.—The first pen deflection of 2 mm. (.67 mm. at the eye) or more occurring within the interval from 200 msec. after CS onset to UCS onset was scored as a CR. The percentages of CRs for the last four trial blocks are presented in the lower row of Fig. 1. Since the scoring interval increases in length as ISI increases, a greater probability of a spontaneous blink falling into the scoring interval and being identified as a CR would be expected as a function of ISI. As a simple correction for differences in the influence of spontaneous blinking as a function of ISI, the first pen deflection of 2 mm. or more occurring within the interval of the same length as the scoring interval but immediately preceding the onset of the CS was scored as a spontaneous blink, and each *S*'s CR frequency was adjusted by subtracting, trial block by trial block, the frequency of spontaneous blinks from the frequency of CRs. The corrected percentages of CRs for the last four trial blocks are presented in the upper row of Fig. 1. It can easily be seen that while this correction procedure has the effect of reducing CR frequencies at the longer ISIs, it does not increase the difference between trace and delay conditioning performance at any ISI.

Each *S*'s record was also scored according to the response derivative criterion as suggested by Hartman and Ross (1961). In Exp. I, the distribution of voluntary form responders for trace and delay conditioning, respectively, was 500 ISI, 4, 4; 800 ISI, 2, 3; 1,100 ISI, 1, 2; 1,400 ISI, 3, 3. In Exp. II the distribution of voluntary form responders was as follows: 250 ISI, 6, 4; 300 ISI, 4, 7; 350 ISI, 8, 5; 500 ISI, 7, 5. The remaining *S*s were classified as conditioned form responders and their CR frequencies, both uncorrected and corrected for spontaneous blinking, are presented in the right column of Fig. 1. It is clear from the figure that the removal of voluntary form responders changes neither the similarity of trace and delay conditioning nor the function relating conditioning and ISI.

Data from each study were analyzed using both the corrected CR frequency and the uncorrected CR frequency for the last four 10-trial blocks. Since the two types of analyses resulted in the same patterns of significant differences, only the results of the analyses of corrected CR frequencies will be presented. An unweighted means analysis of variance for Exp. I, which included ISI, the trace-delay comparison, sex of *S*, and CS frequency as variables, resulted in no significant main effects or interactions. The trace-delay comparison led to $F < 1$.

The analysis of variance for Exp. II included ISI, trace delay, sex of *S*, and CS frequency as variables. Results indicated a significant main effect for ISI, $F(3, 64) = 24.69$, $p < .001$, but not for the trace-delay variable, $F(1, 64) = 1.92$, $p > .20$. A Newman-Keuls test comparing performance at the four ISIs indicated significant ($p < .01$) differences between 250 ISI and all of the other ISIs and between 300 ISI and 500 ISI. The differences between 300 ISI and 350 ISI and between 350 ISI and 500 ISI were not significant.

In addition to the significant ISI effect, there was a significant main effect of CS frequency, $F(1, 64) = 8.04$, $p < .01$, with those conditioned to the 2,100-Hz. tone showing lower levels of performance that those conditioned to the 800-Hz. tone. There was also a significant CS Frequency × ISI interaction, $F(3, 64) = 3.70$, $p < .05$. Comparisons of the two frequencies at each of the four ISIs indicated a significant difference only at 300 ISI, $t(64) = 4.21$, $p < .01$, with the 800-Hz. tone leading to higher levels of performance than the 2,100-Hz. tone. The

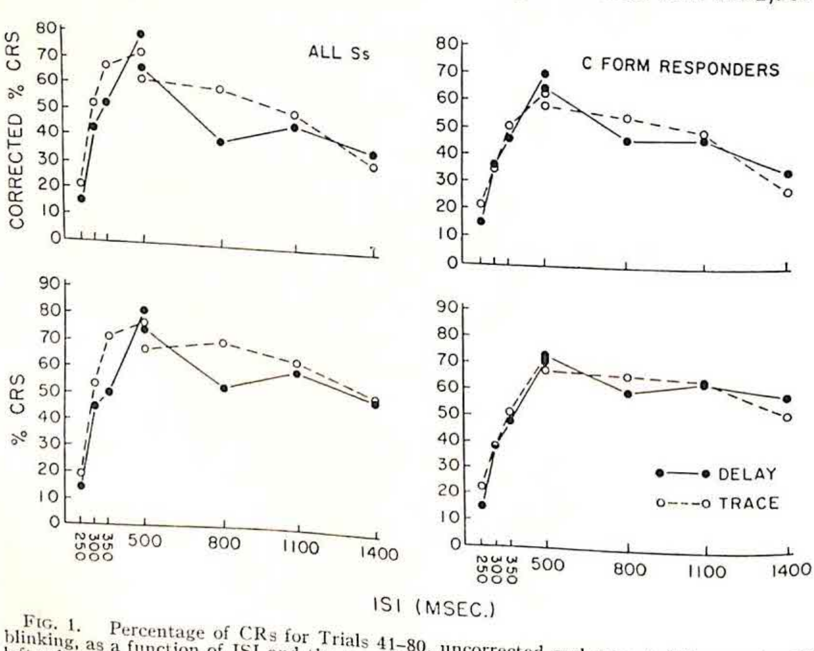

FIG. 1. Percentage of CRs for Trials 41–80, uncorrected and corrected for spontaneous blinking, as a function of ISI and the trace-delay comparison. (Data from all *S*s are in the left column, conditioned form responders only in the right column.)

finding of CS frequency effects is unusual in that the two frequencies, which were matched in loudness, did not lead to performance differences in Exp. I and have not been found to produce different effects in previous studies conducted in this laboratory (e.g., Wilcox & Ross, 1969).

The data clearly demonstrate that this very short trace CS and the delay CS are equally effective in single-cue eyelid conditioning with humans. The comparison of trace and delay procedures at seven ISI values between 250 msec. and 1,400 msec. resulted in no significant differences as a function of CS condition. This finding held for conditioned form responders alone as well as for all *S*s tested. The ISI functions obtained for both trace and delay conditioning are quite similar to those previously found, the highest levels of conditioning appearing within the 400- to 500-msec. range, with lower levels of performance at shorter ISIs and little or no decline in performance at longer ISIs of up to 1,400 msec.

REFERENCES

HARTMAN, T. F., & ROSS, L. E. An alternative criterion for the elimination of "voluntary" responses in eyelid conditioning. *Journal of Experimental Psychology*, 1961, **61**, 334–338.

KIMBLE, G. A. Conditioning as a function of the time between conditioned and unconditioned stimuli. *Journal of Experimental Psychology*, 1947, **37**, 1–15.

MCALLISTER, W. C. Eyelid conditioning as a function of the CS-UCS interval. *Journal of Experimental Psychology*, 1953, **45**, 417–422.

REYNOLDS, B. The acquisition of a trace conditioned response as a function of the magnitude of the stimulus trace. *Journal of Experimental Psychology*, 1945, **35**, 15–30.

SCHNEIDERMAN, N. Interstimulus interval function of the nictitating membrane response of the rabbit under delay versus trace conditioning. *Journal of Comparative and Physiological Psychology*, 1966, **62**, 397–402.

WILCOX, S. M., & ROSS, L. E. Differential classical eyelid conditioning as a function of CS intensity, CS rise time, and interstimulus interval. *Journal of Experimental Psychology*, 1969, **82**, 272–278.

(Received May 17, 1971)

JOURNAL OF EXPERIMENTAL PSYCHOLOGY
MONOGRAPH

Vol. 91, No. 1, 169–190 November 1971

EXPERIMENTS WITH THE STIMULUS SUFFIX EFFECT [1]

JOHN MORTON [2]

M. R. C. Applied Psychology Unit, Cambridge, England

ROBERT G. CROWDER AND HARVEY A. PRUSSIN [3]

Yale University

A number of experiments are described which use the stimulus suffix effect (SSE) to determine the properties of the precategorical acoustic store (PAS) described by R. G. Crowder and J. Morton in 1969. The SSE is a selective impairment of recall of the final items in a serial recall list engendered by a redundant acoustic event following acoustic presentation. First, it is established that such intrinsic properties of the suffix as its meaning, frequency of occurrence, and emotionality have no bearing on the size of the suffix effect. This confirms the precategorical nature of the store. Second, it is shown that variation in the acoustic properties of the suffix (such as its apparent spatial location, timbre, and pitch) with respect to the stimuli reduces the size of the SSE. Variations on this theme indicate that PAS is located after a mechanism of selection between acoustic channels and before the convergence of acoustic and visual analysis systems.

Two of us proposed in an earlier article (Crowder & Morton, 1969) that information concerning the last few items in acoustically presented immediate-memory lists is held in a Precategorical Acoustic Storage (PAS) system. This store is regarded as being a property of that part of the nervous system responsible for the analysis of acoustic inputs. The most obvious consequence of the PAS mechanism is the advantage observed in recall following auditory as opposed to visual presentation, an advantage specific to the last few serial positions. Conrad and Hull (1968) and Routh (1970) showed that if *S* vocalizes rehearsal of visually presented series, the typical "acoustic" result is obtained, while Murray (1965) has shown that such vocalization effects are removed if *S* received white noise sufficient to mask the sound of his own voice. Thus it seems apparent that genuinely acoustic information is implicated in the auditory–visual comparison.

In the earlier article, we demonstrated how the effects of a *stimulus suffix* (Crowder, 1967; Morton, 1968) could be interpreted by reference to PAS. The stimulus suffix (hereafter simply "suffix") is an extra locution pronounced after the terminal element of the to-be-remembered series; its effect is a sharp impairment of the last item or two. Essentially, the suffix effect removes the advantage of auditory over visual presentation. This is both a fair empirical generalization and in addition is the theoretical interpretation advanced by Crowder and Morton (1969). The proposed mechanism for the suffix effect is that the suffix overwrites information in PAS before such information can be used (or in some way "protected") by *S*. It should be noted that the effect of a suffix does not qualitatively resemble the effect of a response prefix (Conrad, 1958, 1960; Crow-

[1] Portions of this research were supported by funds from National Science Foundation Grants GB 4066 and GB 15157. We wish to thank Jean Faulkner, Susan M. Chambers, and Kathleen A. Catanese for their help in collection and tabulation of certain of the data.

[2] Requests for reprints should be sent to John Morton, Applied Psychology Research Unit, 15 Chaucer Road, Cambridge, CB2 2EF, England.

[3] Now at the University of Michigan.

der, 1967). The prefix impairs serial recall at all serial positions, whereas the suffix acts selectively on the very last portion of the function.

The set of questions we are addressing in the present article concerns how the suffix effect is a function of the various acoustic events one chooses to employ as a suffix. After an initial study concerned with the situational generality of the suffix effect, the remainder of the research will be composed of (*a*) studies revealing that the semantic nature of the stimulus suffix makes no difference to the suffix experiment, (*b*) studies showing the rather complex dependence of the suffix effect upon laterality arrangements used, and (*c*) studies exploring the dependence of the suffix effect upon "vocal" or acoustic properties of the suffix event.

The Generality of the Suffix Effect

Experiment I

The purpose of this study was to demonstrate that the suffix effect is not restricted to techniques where complete ordered recall is required. The reasoning was that such a demonstration would support the contention that the results to follow later in this study (virtually all based on the standard ordered-recall method) are of general relevance to human memory and not limited by a particular method of measuring retention. As in many of the studies to follow, the stimuli were lists of digits read at a rate of 2 digits/sec; however, instead of having to recall the whole list from beginning to end, *S*s in the present study were provided with cards showing the series they had just heard in its entirety except for a single digit. It was this single digit, represented by a dash on the test card, that *S* was responsible for recalling. This method constitutes a combination of position and associative probing. The *S* is free to use the item preceding the dash as his cue or he may identify the missing item by noting what position it occupies. In either case, *S* is released from the great burden of (*a*) recalling the entire series and (*b*) recalling order, as well as item, information.

Method.—Fifty-four Yale undergraduates, tested individually, listened to the same fixed list of 108 nine-digit series, each spoken at a 2 digit/sec rate with 1-sec. warning period and three trials a minute presented over a tape recorder. Each *S* served in all three conditions. In the *Control* condition as soon as the list had been presented, *S* was shown a 5 × 8 in. card on which eight of the nine stimulus digits had been typed in a row with a space left for the ninth (tested) digit and a dash drawn in its place. The *S* had before him a numbered list of 108 spaces, in which to write the probed digit for each trial. In the *Prefix* condition, everything was the same except *S* had to speak the word "zero" between hearing the last stimulus digit and receiving the test card from *E*. In the *Suffix* condition, the only change in procedure from the Control condition was that each test series as presented had a tenth element, the word "zero" spoken in the same voice as had originally recorded the series. Six subgroups of nine *S*s each received these three conditions in all possible orderings. However, since the same fixed list of 108 nine-digit stimuli was used for all *S*s, conditions and stimuli were completely balanced against practice and against individual stimuli.

Results.—The results are given in Fig. 1, which shows error probability as a function of condition and serial position. The overall effect of conditions was highly significant, $F\ (2, 106) = 20.03, p < .01$, coming chiefly from the superiority of the Control condition to the two other conditions. The critical outcome was, however, that while the Suffix and Prefix conditions were generally indistinguishable, in the last serial position

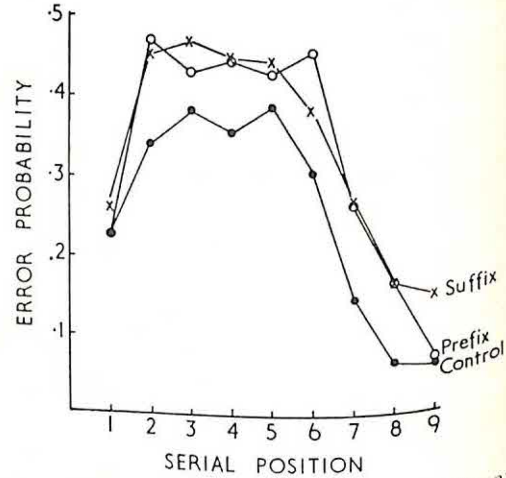

Fig. 1. The relation between error probability and serial position in Exp. I using a complete probe technique where, after presentation of the stimulus, *S* is presented with a response sheet complete save one item.

the Suffix condition revealed over twice as many errors as the Prefix condition. Analysis of the ninth serial position showed that the Prefix condition did not differ from the Control condition but that the Suffix condition did ($p = .002$ by sign test) and that furthermore the difference between the Suffix and Prefix conditions was highly significant ($p < .006$). Elsewhere, there were no significant differences between the two experimental conditions except for a marginal difference at Serial Position 6 ($p = .049$ by sign test) showing more errors in the Prefix condition. We shall assume this latter difference was a chance occurrence.

Discussion.—Exp. I shows simply that the suffix effect, i.e., the selective impairment of recall at terminal serial positions, is readily demonstrable in experimental arrangements other than those involving serial recall. Whether order information is or is not necessary for performance of this probe task is a question beyond the scope of this paper and beyond the grasp of modern theories in any case. However, it seems to us fair to observe that this task does greatly minimize whatever error tendencies lead, in serial recall, to inversions and other order errors. At the very least, the requirements of the present technique and of the standard immediate recall task are substantially different, notably with regard to what *S* must retain about serial order; occurrence of the suffix effect under both circumstances gives some measure of confidence that the phenomenon is not narrowly task specific.

In addition to the standard serial recall task, the suffix effect has been demonstrated with running memory span (Crowder & Morton, 1969) and with complete probing (see Exp. I, above). Still further documentation of the task nonspecificity of the suffix effect is, of course, much to be desired. However, our approach in the remainder of the present paper has been just the opposite. We have largely kept the task the same but have brought about very significant variations in the nature of the memory materials and in the way they have been delivered to our *S*s.

Experiments Confirming That PAS is Presemantic and Precategorical

In earlier experiments, a suffix effect has been found with the digit "zero" following a list of digits (Crowder, 1967; Dallett, 1965) and with an unpredictable digit following a list of letters (Morton, 1968). It is important to know whether there are any semantic effects associated with the suffix phenomenon. Since the underlying model (Crowder & Morton, 1969; Morton, 1969, 1970) specifies that PAS is located well before extraction of meaning, we must predict that there will be no variations in the size of the suffix effect owing to the meaning of the suffix location. The next several experiments confirm this null prediction in a variety of situations.

Experiment II

Method.—The *S*s were 48 volunteers from the Applied Psychology Unit *S* panel, all housewives aged from 25–45. They served in six groups of 8 *S*s each in accordance with a 6 × 6 Latin-square arrangement of the following six conditions: (*a*) *Control* (C): no suffix; (*b*) *Binaural Suffix* (B): the suffix (the word "nought") was at the same apparent loudness and location as the stimulus lists; (*c*) *Monaural Suffix* (M): for half of the *S*s the suffix ("nought") was presented to the left ear and for half to the right ear. The intensity of the suffix was the same as that presented to each ear in Cond. 2; (*d*) *"Recall" Suffix* (R): identical to Cond. 2 except the word "recall" was used rather than "nought"; (*e*) *Random Words* (W): suffixes were randomly selected AA words from the Thorndike-Lorge tables, a new word being used on each trial; (*f*) *Prefix* (P): identical to Cond. 1 except *S* was required to write the digit 0 between hearing the stimulus list and recalling it. A space was provided for this prefix on each line of the answer sheet.

The stimuli were lists of eight digits, binaurally presented. There were 20 lists presented in each block, (i.e., for each of the six conditions) of which the first two were discarded as practice. In addition, *S*s started with a complete practice block in which all the suffix conditions were illustrated. The *S*s were informed before each block of trials which condition would follow. Stimuli were recorded on magnetic tape at a rate of 2 digits/sec and played via a Vortexion tape recorder through S. G. Brown 3C/1100/1 headphones. All suffixes occurred ½ sec. after the final item in a test series (i.e., 500 msec. following the start of the final item).

The stimulus lists were arranged so that no digit was repeated in any list and no digit occurred in successive lists in the same serial position. The digit lists in any condition were arranged so that each digit occurred an equal number of times at each serial position. These precautions ensured that no artifactual influences would contaminate the data.

Recall was written on response sheets with the proper number of spaces. Instructions warned

against starting recall before the list had been completely presented. There were also stipulations calling for ordered recall and for leaving no blanks (i.e., "Guess when uncertain").

In this and several subsequent studies, since the population of *S*s was extremely heterogeneous with regard to memory ability, some method of screening was necessary. Where possible, this was done in advance. On other occasions, *S*s were excluded who either scored the maximum possible number of errors at any one serial position in any condition or who made no errors at all in any one condition. Also, any *S* discovered not to be following instructions (writing down the digits at presentation, recalling in other than serial order, etc.) was eliminated from the study. Groups were then adjusted by random discarding of *S*s from the study so as to balance the control over order of presentation.

Results.—The data from all groups were pooled and the errors at each serial position expressed as the proportion of the number of lists used. Conditions were compared by performing Wilcoxon tests on the pooled data for all serial positions. Only and all of those differences which are significant at better than 2% (two-tailed) are reported.

One might expect that the Binaural "Nought" condition would have a greater effect than either "Recall" or the Random Words conditions, since it belongs to the same set (digits) as the stimuli. In addition, it might be predicted that the Random

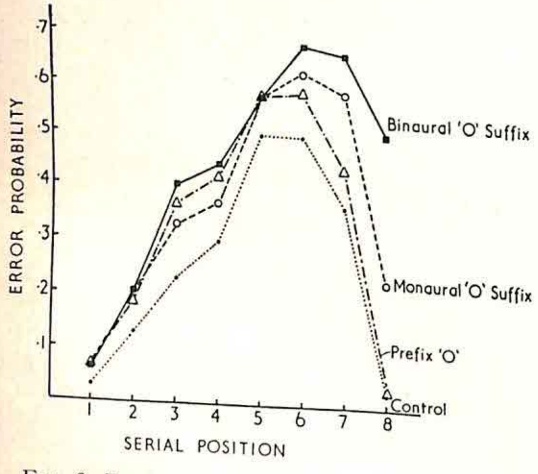

FIG. 2. Error probability in serial recall in Exp. II. (The *Binaural* "0" *Suffix* curve is the average of three conditions indistinguishable in the data. The stimuli were presented binaurally, hence the greater effect of the Binaural Suffix. The difference between suffix and prefix is clear. The prefix and recall were written.)

Words condition would have a greater effect than the "Recall" condition since the latter was completely predictable. Neither of these suppositions are supported by the data, there being no significant differences between these three conditions at any serial position except W and B at Position 5. This isolated difference was attributed to chance. Accordingly, these three conditions are pooled in Fig. 2. All three of these conditions were worse than the Control condition except at Positions 1 (B, W, and R), 2 (R and W), and 5 (B and R). The impairment was greater at Position 8 than at other serial positions and greater at Position 7 than all others except the last. In this respect the Suffix differs from the Prefix, which was worse than the Control condition at Positions 3, 4, 6, and 7, but is indistinguishable from it at Position 8. These results confirm earlier results (Crowder, 1967).

The Monaural Suffix differs from the Control condition at Positions 2, 3, 6, 7, and 8. It was not different from the other suffix conditions except at Position 8. The effects of the monaural suffix will be discussed in detail in later sections of this article.

The important result for the moment is that given both stimuli and suffix are binaurally presented, the effect of the suffix is determined neither by its meaning nor its predictability. The next several studies are directed toward the same point from slightly different angles. We consider it important to furnish considerable redundancy with respect to this result since a null-hypothesis outcome is predicted by the theory.

Experiment III

In this experiment, the prefix and suffix conditions were again contrasted; this time there was a pair of conditions in which the redundant element (prefix or suffix) was the word "zero" and another pair of conditions in which the redundant element was the location "uhh . . . ." This latter event was chosen because it is relatively "impoverished" in an articulatory sense

and contrasts sharply with the articulatory richness of "zero."

Our expectations for these conditions were that while there ought to be no difference between using "zero" or "uhh . . ." as a *suffix*, (for reasons cited above), it might make a difference when these two were employed as prefixes. Crowder (1969b) and Crowder and Morton (1969) have speculated that the prefix effect might result from articulatory interference. (These writers have insisted that the prefix effect is mediated by a different mechanism from that mediating the suffix effect; the nature of the former is not critical to arguments concerning PAS.) If this were the case, then the degree of similarity between the speech gestures involved in *S*s producing the prefix element and those involved in retaining or recalling the memory stimulus should be evident in performance. It seems fair, on this basis, to predict that since the sound "uhh . . ." involves practically no articulatory variety, it should lead to a smaller prefix effect than "zero."

Method.—Each of 20 *S*s (Yale students serving for pay) received the same 100 trials, arranged into five blocks of 20 lists, in the same order. The five conditions, *Control*, *"Zero" Prefix*, *"Uh" Prefix*. *"Zero" Suffix*, and *"Uh" Suffix*, were all presented to each *S*, although in different orders according to a pair of Latin squares. Since the digit lists used on any given trial were the same for all *S*s, the Latin-square arrangement produced complete balance with regard to individual stimuli, and, since the squares used were so chosen, with regard to first-order sequence effects among conditions.

The stimuli were random permutations of the nine digits (excluding zero). A new trial began every 20 sec. and consisted of a "ready" announcement, a 1-sec. pause, then the list of nine digits read at a 2/sec. rate, and finally a silent recall period.

In the two suffix conditions, the extra element ("zero" or "uh") was recorded in *E*'s voice exactly in time with the 2/sec rate. In the prefix and control conditions, only the nine-digit series were recorded. The *S*s were told, before each prefix condition, to emit the prefix as rapidly as possible following stimulus presentation. There seemed to be no problem in getting reasonably standard sounds for "uhh . . ." from the various *S*s; it seemed to many of them an extraordinarily natural thing to do as they prepared to initiate recall. Although recall for the memory series was written, on answer sheets providing the proper number of spaces, prefix emission was vocal. The *S*s were tested individually in a small room with playback through a single loudspeaker.

Results.—Each of the five conditions is plotted separately in Fig. 3. It can be seen in the figure that both types of redundant element had impressive overall effects on error rates whichever word was used. As predicted, more errors were made in the prefix condition when "zero" was used as the prefix element than when "uhh . . ." was used as the prefix element; however, for all positions combined, this result was not statistically significant. Wilcoxon tests showed, in fact, that only at Serial Position 6 were errors in the "Zero" Prefix condition significantly greater than in the "Uh" Prefix condition ($p = .022$).

There was not overall a statistically significant difference between the two suffix conditions, as predicted; however, unexpectedly there was a consistent tendency for the "Uh" Suffix condition to display more errors than the "Zero" Suffix condition in the last few serial positions ($p = .077$ at Position 7, $p = .041$, at Position 8, and $p = .036$ at Position 9). The experiment was, of course, conducted against the possibility that the "Uh" might be the less effective suffix of the two; to have shown a tendency toward the contrary is rather an embarrassment. The best hypothesis at the moment seems to us that acoustic energy reaches *S* earlier chronologically in the "Uh" condition

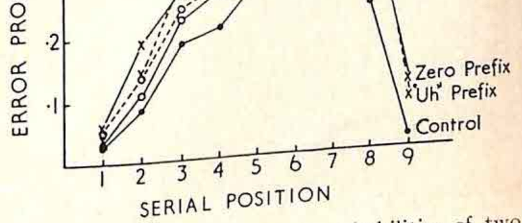

FIG. 3. The effect on error probabilities of two kinds of stimulus suffix and response prefix as a function of serial position. (The prefixes were spoken; recall was written. Data are from Exp. III.)

than in the "Zero" Suffix condition. This difference is inherent in the articulation of the two sounds, "Uh" reaching full intensity immediately and "Zero" not. Crowder (1969a) has recently shown that differences of 100 or 200 msec. in delay of suffix onset after the terminal item appreciably affect the magnitude of the suffix effect. In any case, Exp. III clearly supports the earlier study in failing to find an enhanced suffix effect when the suffix comes from the same semantic class as the memory materials as opposed to when it does not.

Disregarding the "zero–uh" comparison, it is of some interest to compare Fig. 2 and 3, in both of which there is a contrast among prefix, suffix, and control conditions. The obvious and important similarities are that in both studies the prefix and suffix each affect performance adversely, the former nonselectively across serial positions and the latter selectively with regard to items at the end of the list. The one notable difference in the two patterns of results concerns the comparison of suffix and prefix serial position functions. In Exp. II, these two experimental conditions are identical until Position 5, whereas in Exp. III the suffix effect is considerably smaller than the prefix effect until Position 7 where the situation reverses. In an overall analysis of variance in which the two prefix conditions were combined and also the two suffix conditions combined, this crossover showed up in a highly significant Position × Condition interaction, $F(8, 152) = 29.73$, $p < .01$. The most likely explanation for this discrepancy seems to us to be that in Exp. II a written prefix was used, whereas in Exp. III a spoken prefix was used. Crowder and Erdman (1968) have shown that a spoken prefix is more injurious to recall than a written prefix. If the prefix curve of Fig. 2 were simply elevated slightly, the pattern would be identical to that shown in Fig. 3. Other important differences exist between the studies, of course, different *S* populations, different list lengths, etc. The general lesson to be learned from comparing these two studies is not the specifics of procedures or outcomes. Rather, it is that the procedural circumstances of this type of study do make a large difference and that it is therefore not possible to generalize abstractly about what effect these operations have except that the suffix always has a position-specific effect and the prefix always has a nonselective effect. For example, in Fig. 3 if one shifts the suffix condition one position to the left it lines up almost perfectly with the first eight points on the prefix curve; this coincidence invites the conjecture that the prefix and suffix effects are really the same phenomenon but that in the prefix condition an extra event is added at the beginning of the list and in the suffix condition the extra event is added to the end of the list. This proposition is quite opposed to our own theoretical position and to those tempted to make this last interpretation we invite application of the same procedure to the data of Fig. 2 (i.e., shifting the suffix condition one position to the left). In this experiment, the shifted functions are a gross mismatch.

Further indications from Exp. III that the prefix and suffix operations differ fundamentally are provided by consideration of the overall error totals in the four experimental conditions. For every *S*, a number was computed representing the algebraic difference between the "Zero" and the "Uh" Prefix conditions and a second number representing the "Zero" minus the "Uh" Suffix conditions. Under the null hypothesis, that variation in the redundant element affects the prefix and suffix phenomena similarly, these two difference scores should be from the same distribution. However, a Wilcoxon test showed a significant interaction, $T = 56.5$, $p < .038$, indicating that contrasting outcomes in the prefix and suffix situations were observed when the redundant element's identity was varied.

Experiment IV

In this experiment, the critical condition was one in which *S* listened to a series of eight-digit lists with a suffix consisting of the first digit in the list presented again.

The instructions said that this was intended to help them achieve good recall by ensuring that they started right. A naïve associative theory might indeed be shown to predict a simple improvement on this basis. A slightly more sophisticated theory might suppose that the impairment would be worse than with the "nought" suffix since the digit could be regarded as a member of the set of stimulus items. Our prediction was that in agreement with the naïve associative theory, there might be some slight improvement quite early in the list for the repeated *Digit* condition as opposed to the "*Nought*" condition but that thereafter the two suffix conditions would be equally impaired.

Method.—The stimuli were eight digit lists which were presented binaurally. There were three conditions, Control (C), "Nought" Suffix (N), and Digit Suffix (D) (first digit repeated after the eighth), each receiving two blocks of 18 trials in a session consisting of six such blocks. Three groups of 9 *Ss* each were used, with the conditions distributed across the six blocks as follows: CNDNDC, NDCDCN, and DCNCND. In other respects, the procedure for Exp. II was identical to that used in Exp. IV.

Results.—The data from the three groups were pooled and the results are shown in Fig. 4. Wilcoxon tests confirm our predictions. The two suffix conditions differ only at the initial serial position ($p < .01$), the effect of the repeated digit being to give virtually perfect performance. There were smaller differences in favor of Cond. D at the sixth and eight positions, but these were only significant at the 5% level. There was no difference between the total errors made in the two conditions. Thus, the strategy of remaining within the class of digits for choosing suffix conditions but comparing digits which were or were not part of the vocabulary used on any given trial, again, yielded the null result, or, if a 5% significance level is thought worthy of further comment, there was marginal support for a smaller suffix effect from the within-class suffix.

Experiment V

In Exp. II, "nought" was compared with a variety of other words and no difference

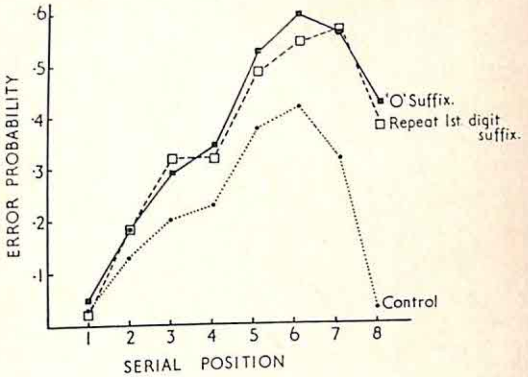

FIG. 4. The effects of a "0" suffix and a suffix which was a repeat of the first digit in the stimulus list (Exp. IV).

was found between the sizes of the suffix effects. It might be argued that "nought" was not properly a member of the digit set and that the effect of the block presentation with the "nought" always in the final position would be to further remove it from the digit set. In addition, it might be argued that in Exp. IV, repeating the first digit introduced complications which also invalidate any comparison on semantic grounds. In the following experiment, we anticipated such objections. The stimulus materials were lists of words from one of two clearly defined semantic classes (animals, utensils), and the suffix events were words either from the same class or from the other class as the to-be-remembered series.

Method.—The stimuli consisted of lists of six words drawn from one of two sets of nine: (*a*) animals: BULL, COW, DOG, HORSE, LAMB, MOUSE, PIG, RABBIT, SHEEP: (*b*) utensils: BOWL, CUP, DISH, FORK, GLASS, KNIFE, JUG, PLATE, SPOON. The materials were chosen to be monosyllabic words and so that no two words in any list began with the same initial letter. There were four blocks of 20 lists each, two of the blocks being made up from the animal list and the other two blocks from the utensil list. Within the blocks there were three suffix conditions which were randomized. For one-third of the lists there was no suffix, for another third of the lists the suffix was the animal name "cat," and for the remainder the suffix was the utensil name "mug." The stimulus population was written up for *Ss* and they were asked to respond by writing the initial letter of the appropriate word. This procedure was welcomed by *Ss*.

It will have been noted that for both stimulus sets, we have suffixes which are either semantically

similar or semantically dissimilar. There were two groups of 10 *S*s; one heard the animal lists first and the other heard the utensil lists first. With the exception that male as well as female *S*s were used in this experiment and also the use of a loudspeaker rather than headphones, the other procedural details were identical to those in Exp. II.

Results.—The data are shown in Fig. 5, pooled so as to reveal the outcome in terms of semantic similarity. There were no differences between the two experimental conditions, both showing the usual suffix effect.

Experiment VI

The preceding experiments seem to indicate fairly convincingly that the semantic relationship between the items in the list and the suffix has no influence on the suffix effect. It remains possible, however, that other properties of the word used as a suffix could influence the data. Thus, while we have established that a semantic code is not involved in the effects which we are interested in, it is still possible that the effects are postcategorical, (i.e., the signal has been uniquely identified), but prior to any semantic look-up procedure. Now there are believed to be certain stimulus properties which influence the categorization process (in the sense in which we use the term). These include the effects of context and the frequency of occurrence of the stimuli (Broadbent, 1967; Morton, 1964, 1968, 1969) and its emotionality (Brown, 1961; Broadbent & Gregory, 1967; Natsoulas, 1965). These properties influence the recognition threshold of the stimuli at the stage which we term categorization. In addition, word frequency is a variable which affects the extent to which the presence of a word as an irrelevant apsect of a stimulus can interfere with color naming the relevant aspect of the stimulus (Klein, 1964). Therefore, if the suffix effect is operating postcategorically, we would expect both the frequency of occurrence and the emotionality of words occurring in the suffix position to exert an influence on the recall data.

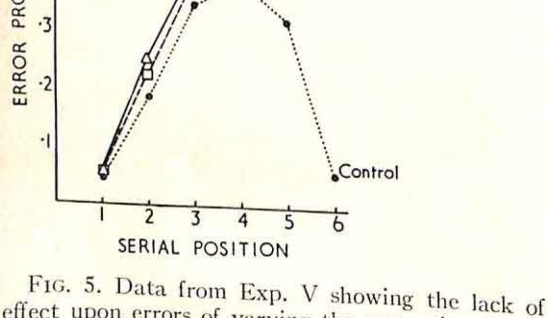

FIG. 5. Data from Exp. V showing the lack of effect upon errors of varying the semantic relation between stimulus items and suffix. (The conditions were randomly applied to lists.)

Method.—The relevant words were taken from Broadbent and Gregory (1967), who provide lists of monosyllabic words of two frequency levels (AA and 10–49 per million) and three classes of emotionality, High, Neutral, and Low. Nine words from each set were selected and the 54 words arranged in a pseudo-random sequence in two blocks of 27. The usual constraints applied to the lists used for each condition both within and between lists and as they appeared in sequence. Each block was preceded by two practice items, and the experiment started with a practice block of 12 lists with two examples of each experimental condition. There were two groups of *S*s who followed exactly the same procedure giving a total of 27 *S*s. Otherwise, the procedure of Exp. V was followed.

Results.—There were no differences to be found between any of the conditions taken singly or between the frequency classes or the emotionality classes. The data are given in Table 1.

Discussion of Experiment XIV

This study will be described in detail below. One of the independent variables it incorporated, however, was a contrast between the words "zero" and "rosy" when used as suffixes in otherwise comparable conditions. Since this comparison is relevent to the questions at hand and quite independent of the other information in Exp. XIV, we will report now that

TABLE 1

SERIAL POSITION ERRORS MADE IN EXPERIMENT VI GIVEN BY CONDITION AND POOLED BY CLASS OF SUFFIX

Cond. and class of suffix	1	2	3	4	5	6	7	8	Total
Good High	12	46	79	117	138	168	163	148	871
Neutral High	13	69	97	108	145	163	173	156	924
Bad High	11	54	99	128	143	174	183	149	941
Good Low	18	62	104	129	129	180	170	151	943
Neutral Low	12	45	78	118	134	173	166	150	876
Bad Low	10	53	91	118	142	174	165	145	898
Total High	36	169	275	353	426	505	519	453	2736
Total Low	40	160	273	365	405	527	501	446	2717
Total Good	30	108	183	246	267	348	333	299	1814
Total Neutral	25	114	175	226	279	336	339	306	1800
Total Bad	21	107	190	246	285	348	348	294	1839

the two conditions were indistinguishable both statistically and visually. We were, of course, aware that these two words are not phonemic reversals of one another (see Crowder & Raeburn, 1970). However, they do not present grossly dissimilar sounds.

Conclusions about Semantic Similarity

There are numerous further experiments one could do on how the suffix effect responds to variation in semantic similarity between the memory series and the suffix. We leave this further research to other investigators, however, being ourselves convinced that there are no such effects to be found. The experiments reported have failed to discover any systematic differences in the size of the suffix effect which might be attributable to intrinsic, as opposed to extrinsic, properties of the suffix. We regard this as strong evidence that PAS is "located" prior to any categorization process in spite of the evidence being all negative. The reason for this will become apparent in sections to follow where the phenomenon is revealed to be very sensitive to variations in the acoustic properties of the suffix.

EFFECTS OF PHYSICAL DIFFERENCES BETWEEN STIMULUS AND SUFFIX

While we did not expect there to be any differences in the suffix effect due to semantic variables, we would expect that changes in the physical nature of the suffix would lead to differences in recall of the final items in a list. We had, however, no prior expectations as to exactly which of the possible variables would have an effect and which would not, except a general belief that if a dimension proved to be important, then we would be able to describe the phenomena in terms of the similarity of the suffix to the stimuli. Thus far our notion of PAS is restricted to the belief that it is a property of some section of the acoustic analysis system, but it has not been necessary for us to specify where. As long as PAS is precategorical and is specifically acoustic, then, as far as its position in terms of current theories of memory is concerned, it is secure. The existence of the suffix effect enables us to be more precise, for it is undoubtedly of interest whether or not we are dealing with a phenomenon specific to the basilar membrane, one which exists centrally but separately for the two ears, or whether inputs to the two ears are treated alike in this respect.

Effects of the Laterality of the Suffix

What the following series of experiments demonstrates is that the information contributing to the effect with which we started (the difference between auditory and visual presentation on the final items) is located in a part of the system after the combination of information from the two ears, at a stage where intensity has been normalized but where a number of other variables are still separable.

In Exp. II, we reported comparing a monaural suffix with a binaural suffix following binaural presentation of the stimulus lists. The difference between these two conditions was restricted to the last serial position in the list. At this position, the Monaural Suffix condition yielded significantly more errors than the no-suffix conditions (see Fig. 2). This effect was, however, significantly smaller than that found with the Binaural Suffix condition. From this result, it may be concluded that there are two effects of the suffix, one of them independent of the location of the suffix and affecting several of the terminal serial positions (including the last position). The other consequence of the suffix varies in strength with the location in auditory space of the suffix element and is restricted to the final list position. We shall return to this distinction in the later sections of this paper.

Experiment VII

The present investigation was directed at a simple and straightforward question: Given the memory series has been received in only one ear, what differences are there, if any, between then receiving a suffix in that same ear as opposed to receiving the suffix in the other ear?

Method.—Six groups of four *S*s each listened to six blocks of 27 stimuli lists, each block preceded by 2 practice lists and the whole session preceded by 18 practice lists. There were eight digits in each list. The stimulus items were recorded on one channel of a magnetic tape and the suffix "nought" was recorded on the second channel. All groups heard the lists in the same order, six conditions being created by means of external switching gear. The suffixes were at the same pitch and loudness as the stimuli. The six conditions were as follows: (*a*) LC: the stimulus lists were presented in the left ear; there was no suffix; (*b*) LL: both the stimulus and the suffix were presented to the left ear; (*c*) LR: the stimuli were presented to the left ear and the suffix to the right ear; (*d*) RC: the stimuli were presented to the right ear; no suffix; (*e*) RR: both stimuli and suffix were presented to the right ear; (*f*) RL: the stimuli were presented to the right ear and the suffix to the left ear. Conditions LC and RC were control conditions; RR and LL will be referred to as the "ipsilateral" (Ipsi) conditions and RL and LR as the "contralateral" (Contra)conditions.

The six groups experienced the six conditions according to a Latin-square design. The initial practice items demonstrated all the alternative conditions, and the two practice items at the beginning of each block were of the same form as the block. All *S*s were right-handed.

In other details, the procedure of this study followed that of Exp. II.

Results.—All conditions were compared with all other conditions for errors at each serial position and total errors using the Wilcoxon test. There were no differences at the 5% (one-tailed) level between the two control conditions, between the two ipsilateral conditions, nor between the two contralateral conditions at any serial positions. Accordingly, these pairs of conditions were combined for presentation. The resulting data are shown in Fig. 6. The following differences were significant at 1% or better (two-tailed) comparing only conditions with the stimuli in the same ear: Control versus Ipsi, Serial Positions 3–8; Control versus Contra, Serial Positions 1 and 5–8; Ipsi versus Contra, Serial Positions 7 and 8. Examination of the interactions between stimulus ear and condition revealed that the only reliable differences (statistically) were that the contralateral suffix had a greater effect on the left ear stimuli at Positions 1 and 6, ($p < .01$, Wilcoxon, two-tailed). It does not seem that such data justify any claim to have found laterality effects considering the number of comparisons which were made and lack of overall difference between the two ears.

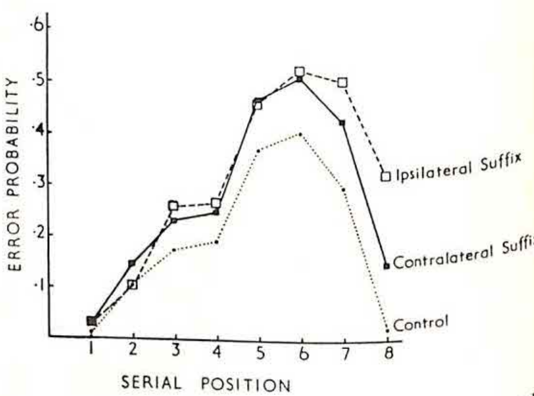

FIG. 6. The effects of a suffix in the same ear as the stimulus list (Ipsilateral) and in the opposite ear (Contralateral) compared with a control condition with no suffix (Exp. VII).

In addition to the general effects of the suffix referred to above, the behavior of the contralateral suffix is similar to that of the monaural suffix following binaural presentation in Exp. II. The lack of any interesting differences between the two ears is not too surprising as the situations in which such effects have been discovered have involved very careful control of the timing of stimuli in the two ears (Kimura, 1961; Shankweiler & Studdert-Kennedy, 1967, 1970).

Discussion

Our ideas concerning the specific nature of PAS have, up to now, been restricted to supposing that it is a consequence of the nature of some part of the nervous system whose function is to analyze acoustic signals and transform such inputs to a form in which they can be processed by mechanisms which can also accept inputs from other sources. This is the force of the "acoustic" part of PAS. The experiments reported above on the effects of the laterality suggest two broad possibilities.

We can imagine, to start with, that the Acoustic Analysis System comprises three parts, one specific to each ear and one common to the two ears. Information concerning the final item(s) of an acoustically presented list under conditions of serial recall could possibly be obtained from any or all of these three parts. If the stimulus were monaural, then two of the sections would be involved in its analysis—that part specific to the stimulated ear and the part common to the ears. This state is indicated in Fig. 7a. An ipsilateral suffix would be processed by the same two parts and would thus affect all the available PAS information. The result would be the typical "visual" curve—with very small recency effects—as shown in the top, Binaural graph in Fig. 6. A contralateral suffix, on the other hand, would only affect that part of the system which was common to the two ears, leaving the ear-specific part of the system intact; as indicated in Fig. 7b. We would thus expect that the contralateral suffix would produce a decrement compared with the control condition (owing to the loss of information in the common part) but would not have as great an effect as the ispilateral suffix (since the information in the ear-

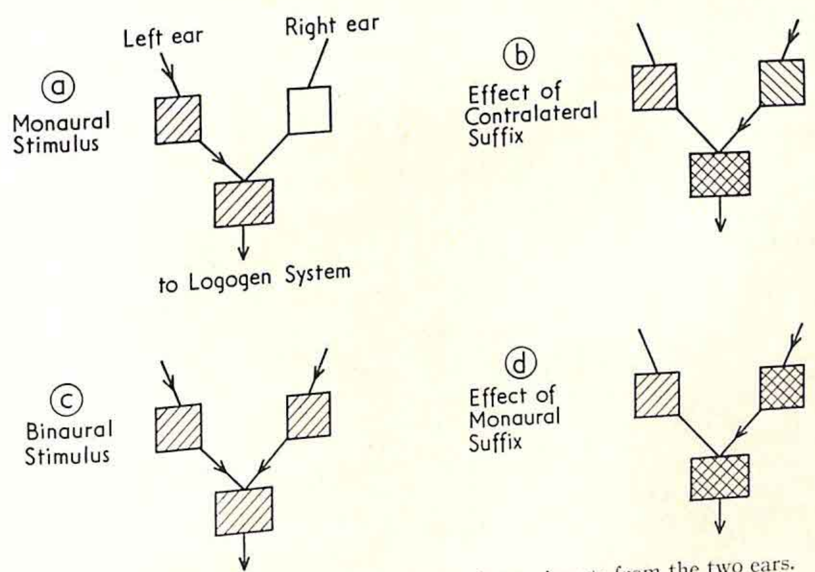

FIG. 7. One possible model of PAS in relation to inputs from the two ears. (The boxes correspond to "stores," one specific for each of the ears and one common to the ears. A stimulus leaves a trace in the appropriate stores as indicated in a and c for monaural and binaural stimuli, respectively. The suffix has the effect of erasing the trace in any store it passes through; this is shown by the cross-hatching in b and d. This accounts for the data of Exp. II and VII. This model predicts that a binaural suffix would eliminate the traces of a monaural stimulus, having as much effect as an ipsilateral (same ear) suffix. This prediction was falsified in Exp. VIII.)

specific part of the system would be unaffected). This would account comfortably for the data shown in Fig. 6.

With binaural presentation, all three parts of the system would be involved as shown in Fig. 7c. A binaural suffix would then affect all three parts leaving no information in PAS. A monaural suffix, however, would only affect that part of the system which was specific to the ear being stimulated together with the common part, leaving the other, ear-specific part unaffected. This is shown in Fig. 7d. Thus we would obtain the result described in Exp. II. Chronologically, Exp. VII was performed before Exp. II, the latter being, in part, a test of the model just described (see Morton, 1970).

One condition which had not, at the time, been presented, was the case of a monaural stimulus followed by a binaural suffix.[4] In the model described above, such a suffix would have the effect of eliminating all the PAS information since it would be equivalent to an ipsilateral suffix plus additional stimulation in the other ear. Thus we would expect a binaural suffix to have the same effect as an ipsilateral suffix, recall of the last item for both of these conditions being inferior to that with contralateral suffix.

On an alternative model of the process, a totally different prediction can be made. This is a model based on Broadbent's (1958) filter model and by analogy with the results on selective listening (Treisman, 1964a, 1964b). In such models, there is a point in the processing of stimulus information where particular channels (specified, for example, by spatial location) may be accepted or rejected. In terms of a filter model, PAS information may be located either before or after the filter. In either case, however, the model would predict that a binaural suffix would have less effect following a monaural stimulus than would an ipsilateral suffix since *binaural* specifies a different channel from *monaural*. PAS could be located before the filter, since the binaural suffix would, in some sense, be stored separately from the monaural stimulus and would not interfere with it. If PAS information is retrieved from after the selection mechanism, then we can assume that a suffix which is not ipsilateral can be selected out. The comparison of a binaural suffix with ipsilateral and contralateral suffixes following monaural presentation then becomes a crucial test between these two models of the processs.

[4] We are grateful to D. W. J. Corcoran for pointing out this rather obvious omission.

Experiment VIII

This study was designed to compare the effects of a binaural suffix and those of an ipsilateral suffix following monaural presentation of the stimulus list.

Method.—Three conditions were employed, in all of which the stimuli were played to the right ear. The stimuli, six blocks of 18 lists of eight digits each, were recorded on one channel of a two-channel Vortexion tape recorder, and the suffix "nought" was recorded on the second channel. External switching and mixing gear were used to provide the following three conditions: (*a*) *Ipsilateral Suffix* (I)—the intensity of the suffix was adjusted so that there was no perceptible change of loudness between the stimuli and the suffix. (*b*) *Contralateral Suffix* (C)—the suffix was presented to the left ear at exactly the same intensity as in Cond. I. (*c*) *Binaural Suffix* (B)—the suffix was presented to both ears, the intensity in each ear being exactly the same as in the above conditions. This meant that the loudness of the binaural suffix was greater than the loudness of the other suffixes.

There were three goups of four *S*s each. Each group listened to the blocks of stimuli in the same order but the suffixes were presented according to two 3 × 3 Latin squares. Otherwise, the procedure was the same as that used in Exp. II.

Results.—The data are presented in Fig. 8. In this graph, the data from all the groups have been pooled. The control curve, included for guidance only, is the mean of the control curves from other ex-

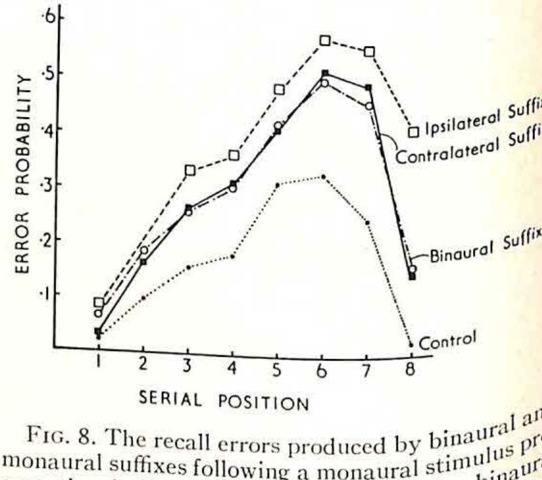

FIG. 8. The recall errors produced by binaural and monaural suffixes following a monaural stimulus presentation in Exp. VIII. (The effect of a binaural suffix being less than that of an ipsilateral suffix falsifies the model in Fig. 7. The control curve is an average from other experiments, being included for guidance only.)

periments, encompassing over 4,000 observations from 99 *S*s. Wilcoxon tests revealed no significant differences between Cond. C and B even at the 5%, two-tailed level, at any serial position. Ipsilateral Cond. I was significantly worse than the other two conditions on the final serial position and overall ($p < .01$, two-tailed) and also significantly worse than the B condition at Position 7.

This result means that the first of the models proposed above (and in Morton, 1970) must be rejected, and we must conclude that PAS phenomena are due to processes which operate after the combination of information from the two ears, though, of course, the "channels" remain separate. It might be noted that the result of this experiment constitutes a very strong falsification of the rejected hypothesis. If the binaural suffix had produced a greater effect than the contralateral suffix, it could have been argued that this was because the binaural suffix was the louder, and as such more difficult to reject at the filter. This is the result one would have expected by analogy with the results in selective listening (Treisman, 1964a). As the binaural suffix—in spite of its greater loudness—did not have a greater effect than the contralateral suffix, we have three possible assumptions in this experimental situation with regard to the processes contributing to PAS: (*a*) that the loudness of a stimulus is ignored and so must be coded in digital fashion prior to the location of PAS, the stimulus iteslf being thus normalized; (*b*) that a binaural stimulus is "tagged" as having an intensity appropriate to one of the ears only. In this case, the information which gives rise to a subjective impression of increased loudness must be separated from the coded suffix; (*c*) that loudness is itself a cue upon which selection or rejection can be based. If this were not the case, then the binaural suffix might be having *less* effect than if it were equal in loudness to the stimuli.

Experiment IX

The purpose of this study was to check against some of the possible artifacts mentioned above. In particular, the role of intensity was examined in a design otherwise identical to Exp. VIII.

Method.—The procedure, materials, and conditions were all identical to those used in Exp. VIII except that there were the following six suffix conditions: (*a*) *Ipsilateral* (I)—suffix same loudness as the stimuli; (*b*) *Ipsilateral+* (I+)—suffix presented at 3.75 db. greater intensity than in I; (*c*) *Contralateral* (C)—suffix in opposite ear as stimuli at same intensity as I; (*d*) *Binaural* (B)—suffix presented to both ears at same intensity as in I and C; (*e*) *Binaural−* (B−)—suffix presented at intensity 3.25 db. lower than in Cond. B. At this level, the suffix was judged as being at about the same loudness level as the stimuli; (*f*) *Binaural+* (B+)—suffix presented at 3.75 db. above the level in Cond. B.

The changes in intensity were achieved by means of a gain control on the tape recorder, the accuracy of settings being better than ± .25 db. Six groups of four *S*s listened to the stimuli with the suffixes being presented according to a 6 × 6 Latin square design.

Results.—The data are shown in Fig. 9. There were no statistically significant differences at the 1% level in the number of errors made between the two I conditions, among the three B conditions, or between

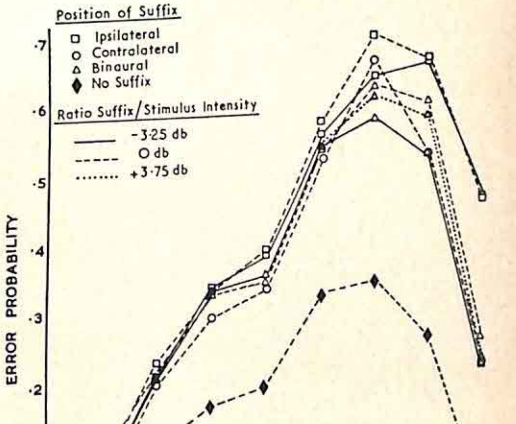

FIG. 9. An examination of the interaction of the apparent spatial location of the suffix and its intensity. (Two levels of intensity of the ipsilateral suffix were used (0 db. and 3.75 db. with respect to the stimuli. Three intensity levels were used for the binaural suffix (−3.25 db., 0 db., and 3.75 db.) and one (0 db.) for the contralateral suffix. There were no reliable effects of intensity (Exp. IX). The control curve is taken from other experiments.)

TABLE 2

RESULTS OF WILCOXON TESTS IN EXPERIMENT IX

Cond.	Cond.			
	B −	B	B +	C
I	7, 8, T.	8	8, T.	7, 8, T.
I +	6, 7, 8, T.	8	6, 8, T.	7, 8, T.

Note.—The numbers indicate the serial positions at which the I conditions showed more errors than the other conditions ($p < .01$, one-tailed). T refers to total errors.

the B conditions and C except for a slight increase in the number of errors on the initial position in Cond. B (B vs. C being the only difference which reached significance). The two I conditions differed from the other conditions as shown in Table 2. These data confirm the results of Exp. VIII and also indicate that the loudness of the suffix with respect to the stimulus is not a factor of importance in the suffix effect. (This latter conclusion will be discussed further below.) There is an interaction effect apparent in significance levels given in Table 2 such that the advantage of the B condition over the I and I + conditions is less than that of the B + and B − conditions for total errors and Serial Positions 6 and 7. This indicates that the binaural suffix has a *greater* effect

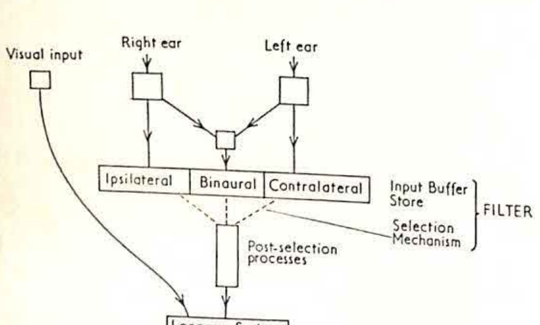

FIG. 10. A possible flow diagram of information in the case where the stimulus list is presented monaurally to the right ear. (The *logogen system* (described elsewhere Morton, 1969, 1970) is the point at which information from visual and acoustic stimulation converges. Since there was no visual-acoustic interactions with suffix experiments (Crowder & Morton, 1969; Morton & Holloway, 1970) PAS must precede the logogen system. The possible location of PAS is discussed in the text.)

when its intensity in each ear is the same as that of the stimulus list as was the case in Exp. VIII. This is a small effect, however (as there are no significant differences at all between B and B + or B and B −), and one which does not find any equivalent with the ipsilateral suffix. It might then be that intensity is a cue whose effect is too small to be seen in isolation but which interacts with another cue (laterality) to produce a visible effect. Further experiments will be required to decide this issue. It is quite clear, however, that loudness is not a significant factor on PAS either as a selection cue (in which case we would have expected B and B +, which were louder than the stimuli, to have less effect than B − or C), or as an attention-getting variable (in which case we would expect B + to have more effect than B or B −).

Discussion.—Having concluded that PAS is a property of processes which follow the integration of information from the two ears, we now have to decide, in terms of the model shown in Fig. 10, whether PAS information is located before or after any selection mechanism. The two possibilities may be summarized as follows.

1. PAS information is located in the input buffer store. If the buffer store segregates information from different locations, then while an ipsilateral suffix overwrites PAS information, a suffix with a different spatial origin, i.e., from a different channel, will leave PAS undisturbed. This would account for the difference between the ipsilateral and the other suffixes. The fact that there is a suffix effect in the B and C conditions would be a result of incomplete separation of stimuli from different sources in the buffer store.

2. PAS information is located in the post-selection processes. The difference between the ipsilateral and the other suffixes would then be due to the selection system being set for the laterality of the stimulus list and subsequent items on another channel being "attenuated" (cf. Treisman, 1960). The extent of the suffix effect with suffixes on other channels would then be an index of the efficiency of the selection mechanism.

3. PAS information might be located both in the buffer store and in the post selection processes. In this case, the above restrictions would operate together.

Experiment X

In the experiments described above, *S*s always knew from whence the suffix would come and it can reasonably be assumed that they adopted a strategy to minimize its effect. It would thus be a further assumption of the above discussion that it is easier to exclude a suffix on a different channel from the stimulus (by maintaining the current filter setting) than to exclude a suffix on the same channel (by switching the selector to some other channel or trying to exclude *any* input).

If PAS is a property of the postselection processes, then it should be possible to increase the effects of a binaural or a contralateral suffix by keeping *S* uncertain as to the channel on which the suffix will be presented (assuming that the location of a rejected channel must be specified) or by forcing *S* to process the suffix before starting to recall the stimuli. In both these cases, the difference between the ipsilateral suffix and the others should be reduced. If, on the other hand, PAS is entirely a property of the buffer store, then such experimental manipulations should make no difference to the comparison between the different suffixes. In the present experiment, the suffix occurred at random on the right ear, the left ear, or binaurally.

Method.—The stimuli, lists of eight digits, were presented to the right ear and the intensities of the suffixes were determined as in Exp. VIII. There were four blocks of 29 trials of which the first two were not scored. The test stimuli were preceded by 16 practice trials for which the channel of the suffix was also randomized. There were 17 *S*s. In other respects, the procedure was the same as that used for Exp. II.

Results.—The averaged data are shown in Fig. 11 together with control data which were taken from several other experiments. Differences among conditions were tested with the Wilcoxon test. The Binaural and Contralateral conditions were indistinguishable statistically. The Contralateral Suffix condition differed significantly from the Ipsilateral condition at Positions 2, 5, and 8 and also for total errors ($p < .01$). The Binaural condition, however, was not significantly different from the Ipsilateral

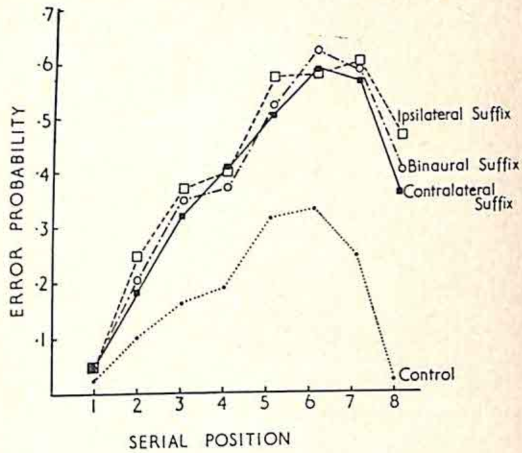

Fig. 11. The effect on errors of suffixes occurring in three spatial locations at random (Exp. X). (The differences between the Ipsi condition and the other two are reduced compared with Exp. VIII in which *S*s always knew the suffix location (cf Fig. 8). The control curve is taken from other experiments.)

condition at this level, the difference between these reaching only the 5% level at the last serial position (two-tailed). On more detailed examination, this turned out to be a consequence of two *S*s' performance, and a sign test showed the Binaural condition to be sharply better than the Ipsilateral condition for the last position ($p < .004$). Thus, we are not prepared to conclude that there is not a difference between binaural and ipsilateral suffixes under the conditions of this experiment.

The Effects of Random Suffix Positions

A comparison of the data in Exp. VIII and X (Fig. 8 and 11) will enable us to draw some conclusions as to the effect of the spatial location of the suffix being unpredictable. There are two effects visible: an overall increase in the number of errors and a reduction in the difference between the Ipsi and the other two conditions. As there are wide differences in ability between *S*s within each experiment, Mann-Whitney tests are rather weak, but confirm the visual impression, yielding significant differences between the experiments in the final serial position for both the number of errors in the B and C conditions ($p < .01$) and for the I–B and I–C differences ($p < .05$).

From this result, we conclude that in terms of the filter model, at least a large part of the PAS information is located after the selection mechanism. Otherwise, we would not expect the suffixes on different channels to the stimulus to have a greater effect as a result of unpredictability in suffix positions. In terms of the average error rate, there is also a suspicion that the number of errors made on the ipsilateral condition was greater in Exp. X. This difference could not be tested adequately owing to extent of the inter-*S* variance (from 24 to 186 total errors for individual *S*s in Exp. X and from 31 to 179 in Exp. VIII).

Effects of Suffix-Prefix

In the next two studies comparisions were again arranged among ipsilateral, binaural, and contralateral suffixes. The difference in procedures from Exp. VII–X was that in Exp. XII the suffix had to be *processed* by *S*. The increase in information-handling requirements by *S*s required moving to lists of seven digits rather than eight; otherwise, too many errors were made in recall. The suffix elements used were either the word "tick" or the word "cross," the choice for any given seven-digit series being determined randomly. In Exp. XI and XII, the suffix occurred at a random spatial location (ipsilateral, contralateral, or binaural) throughout. In Exp. XI, the suffix had to be ignored; in Exp. XII, *S* was asked to respond appropriately to the suffix before beginning recall (i.e., to write either a tick or a cross on his answer sheet before writing the digits). In spite of the loss of statistical power, we did not feel that we could use the same *S*s in both experiments owing to the danger of asymmetric transfer effects. In both experiments there were four blocks of 29 items preceded by 16 practice lists.

Experiment XI

Method.—The procedures and materials were similar to those used in Exp. X above. The three conditions were Binaural Suffix, Ipsilateral Suffix, and Contralateral Suffix; 20 *S*s were tested in four groups.

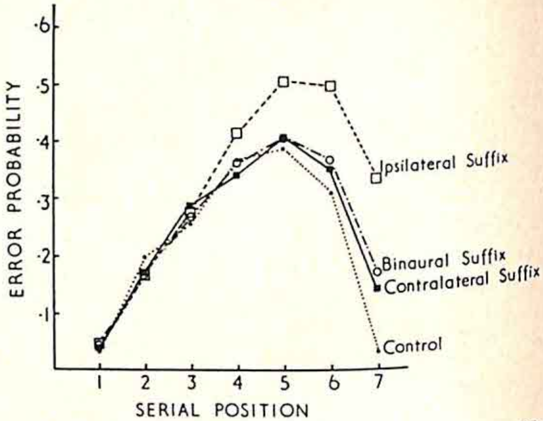

FIG. 12. The errors found in Exp. XI due to a suffix of the words "Tick" or "Cross." (These data confirm those of Exp. X (Fig. 11) and act as a control for Exp. XII (Fig. 13). The control data are for guidance only, being taken from other experiments.

Results.—The error probabilities in these three conditions are shown in Fig. 12. The control data were taken from other experiments and are there for guidance only. Wilcoxon tests permitted the inference that performance in the Ipsilateral condition was significantly worse than in the other two experimental conditions on Serial Positions 5, 6, and 7, as well as for total errors ($p < .01$, two-tailed). There were no significant differences between the Binaural and Contralateral conditions. When this outcome is compared with the data of Exp. VIII (Fig. 8), an interaction

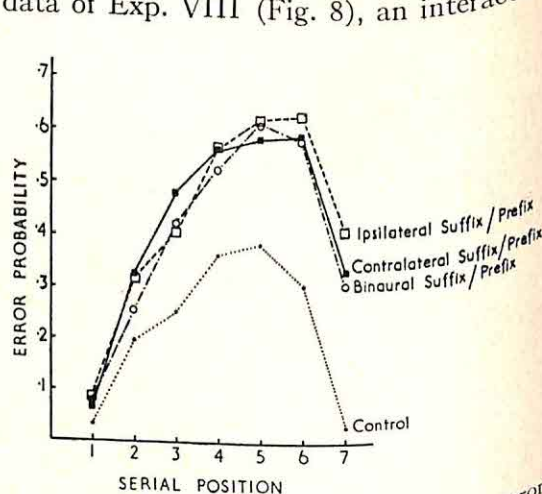

FIG. 13. Data from Exp. XII showing the errors caused by a suffix ("Tick" or "Cross") which had to be responded to as a prefix before recall of the stimuli. (The control curve was taken from other experiments).

is suggested between list length and the differential effects of the suffix such that with the shorter lists (and the same population of *S*s, though not the same individuals) the difference between the Ipsilateral condition and the other two is greater on the final serial positions than with the longer lists used in Exp. VIII. Comparison of the size of the effects of the ipsilateral suffix with different list lengths would be impracticable without a much more detailed model of all the processes involved in the recall task.

Experiment XII

Method.—All details were identical to those in the preceding experiment except that *S* was required to respond with the suffix (either "tick" or "cross") as a prefix before recalling the digits. Seventeen *S*s were tested.

Results.—The error probability scores are presented in Fig. 13.

The only significant difference among conditions was at Position 3, where performance was better following the ipsilateral suffix than following the contralateral suffix, and at Position 7, where performance after the ipsilateral suffix was worse than that following the binaural suffix. This result suggests the hypothesis that when *S* is required to process the suffix the difference between channels is reduced. This impression is confirmed by the results of a Mann-Whitney test shown in Table 3, which shows those serial positions at which the difference between conditions is reduced. This table shows that the differences between Ipsi and the other conditions are reduced more at the later serial positions. This result indicates a relative increase in the suffix effect of the Binaural and Contralateral Suffixes. It should be noted that the tests of differences between Exp. XI and XII are highly conservative. If *ratios* of error probability were used instead of differences, the serial position effect of the change in procedure would be emphasized more. Even without such a test, we feel confident that there is sufficient evidence to justify the claim that the comparison of Exp. XI and XII supports the conclusions from the comparison of Exp. VIII and X that PAS information is located after the channel selection. Only in this way can one account for a differential effect of forcing *S*s to process the suffix. The overall error rate is higher in Exp. XII than in Exp. XI as a consequence of the added Prefix operation.

Effects of Noise and Vocal Differences

Experiment XIII

We have shown in the first section that in a large number of situations the semantic properties of the stimulus suffix make no difference to the degree of performance failure occasioned by suffix persentation. In the section just concluded we have shown that spatially defined channel separation of the stimulus from the suffix can reduce the magnitude of the effect. It now remains to explore how physical properties of the suffix other than spatial location can influence the suffix effect.

Experiment XIII was designed to test our early conjecture (which now seems naïve) that nearly *any* sound would produce a suffix effect; this conjecture presupposes, of course, that PAS is more peripheral than several of the above studies have given cause to believe. Control series of nine

TABLE 3

VALUES OF MANN-WHITNEY *U* STATISTIC: COMPARING EXPERIMENTS XI AND XII

Difference compared	Serial position						
	1	2	3	4	5	6	7
Ipsilateral–Contralateral	135	137	108.5	88**	94.5*	65***	92**
Ipsilateral–Binaural	167.5	112	124.5	152.5	68***	96*	113

* $p < .05$
** $p < .02$
*** $p < .002$

digits were compared with four experimental conditions. The latter were arranged as a 2 × 2 factorial design; one factor was whether the suffix event was a spoken "zero" or a burst of random speech noise and the second factor was the intensity of the suffix.

Method.—Ten *S*s served for two sessions on different days, each session consisting of five blocks of 20 trials. Conditions (Control, Soft "Zero." Loud "Zero," Soft Noise, Loud Noise) were assigned to blocks according to two Latin squares (one for each day) such that first-order sequence effects between conditions were completely balanced. Two lists of 100 nine-digit stimuli were used, one list being used on Day 1 for half of the *S*s and the other list on Day 2 for the remaining *S*s.

All stimuli were read at a 2 digits/sec rate with the suffix arriving in time with the memory series. In the Control condition, no suffix was presented. In the two "Zero" Suffix conditions, *E* recorded the word "zero" following the last memory element such that in the Soft condition the intensity was the same as for the digit list and such that in the Loud condition the suffix intensity was subjectively twice as loud as the digit series. Subsequent measurement showed peak intensities of the Soft "Zero" to be the same as the memory series (around 66 db.) and the Loud "Zero" to be at around 85 db. In the noise conditions, a burst of white noise (actually a broadband noise signal designed to cover approximately the same bandwidth as average human speech) was presented at the same time and for the same duration as the "Zero" Suffix. Subsequent measurement of of these noise signals showed the Soft Noise to be 80 db. and the Loud Noise to be 90 db. Thus, it should be remarked that according to SPL readings the design of this study was not factorially pure—both noise signals were substantially more intense than the normal speech conditions. However, we felt it more important that the subjective loudnesses be matched with some care. In calibrating the noise suffixes, the subjective criterion applied was that the soft noise should sound equally loud as the intensity of the to-be-remembered digits (and hence equally loud as the Soft "Zero" condition) and that the Loud Noise be twice as loud.

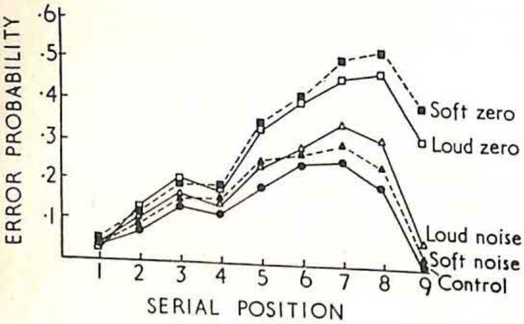

FIG. 14. Data from Exp. XIII showing that a nonspeech noise gives no suffix effect. (The difference between "Soft Zero" (where the suffix was of equal loudness with the stimuli) and "Loud Zero" is properly to be attributed to differences in voicing not intensity.)

Results.—The results are shown in Fig. 14. The most striking finding is that while both conditions in which a speech response was used as the suffix caused a suffix effect, there was none whatever when the nonspeech noise was used. Overall, the two noise conditions (which did not differ from one another either in total errors or in errors on Positions 7 and 8) occasioned more errors than the control condition, $T = 7$, $p < .019$; however, it is obvious from the figure that this was not a *selective* impairment. If the noises affected PAS, we would expect a large decrement on the final serial positions.

The data of Fig. 14 provide an empirical contradiction with Exp. IX in that they appear to show that the Soft "Zero" was a more effective suffix than the Loud "Zero". For Positions 7, 8, and 9 combined, the difference between the two speech suffix conditions was marginally nonsignificant ($p < .055$ by sign test); however, on the last serial position, the difference reached satisfactory levels of confidence, $T = 9.5$, $p < .042$ and $p < .011$, by sign test. We now believe this result is an artifact of the method used in the present experiment. In particular, *E* effected the variation in suffix intensity here by simply speaking louder (nearly shouting in fact) when he was recording the Loud "Zero" condition than when he was recording the Soft "Zero" condition. This shouting procedure probably entailed variation in much more than intensity alone—pitch differences, constriction of the vocal apparatus, stress differences, etc. As we shall describe below in some detail, our subsequent work has indicated these factors to be of large probable importance in the suffix experiment. Finally, since intensity alone in Exp. IX (where it was properly varied by adjustment of the gain control on playback equipment) was not an effective determinant of the suffix effect, we are confident in attributing the present difference to

factors confounded with the present manipulation of intensity.

Experiment XIV

In the preceding experiment, we have demonstrated that the set of sounds which will produce a suffix effect does not include bursts of noise. Knowing from the series of studies described above that laterality is a factor in the suffix effect and knowing also from the first group of experiments that semantic factors do not affect the suffix experiment, a correspondence with research on selective attention and shadowing is brought into sharp relief. It had been established in such experiments that when the rejected message is in a different voice from the shadowed message, the number of errors made in shadowing is greatly reduced. (Treisman, 1964a). Accordingly, we might expect that when the suffix is in a different voice from the memory series, it too will have less effect. The logic of this prediction is quite clear if one considers that S's task in the suffix experiment is to *ignore* the suffix (an instruction we occasionally use).

Method.—The method in Exp. XIV was exactly the same as that used in Exp. XIII. The five conditions were formed by the Control condition (no suffix) and a 2×2 factorial combination of suffix word (either "zero" or "rosy"—see discussion following Exp. VI above) and suffix voice. The suffix voice was varied as follows. The to-be-recalled digit lists were always read in a male voice; in the Same Voice condition the suffix also was presented in the same male voice, but in the Different Voice condition the suffix was presented in a female voice. Ten Ss were used.

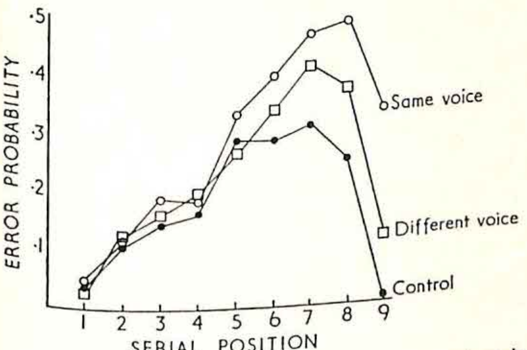

FIG. 15. The effect on errors of a suffix spoken in the same (male) voice and a different (female) voice as the stimuli. (In the control condition there was no suffix—Exp. XIV.)

Results.—Figure 15 displays the main data in terms of error frequencies for the conditions collapsed across suffix word (since, as was observed above, the latter variable had no effect whatever). The major findings were that (*a*) both suffix conditions produced significantly more errors than the control condition, $T = 0$, $p < .005$, for the same-voice suffix and, $T = 10$, $p = .042$, for the different-voice suffix, and (*b*) the same-voice suffix was significantly more detrimental to recall than the different-voice suffix, $T = 0$, $p < .005$. Although these inferences are based on total errors, it is obvious from the figure that the effects are occurring at the last few serial positions. This is confirmed by an analysis of the separate serial positions, the differences at Position 9 being greater than those at other positions for all comparisons. Thus, a shift in voice quality between the stimulus channel and the suffix channel produces a result quite comparable to the shifts in laterality channel studied in several of the experiments above. Both yield an attenuated suffix effect.

Experiment XV

This study was designed to complement Exp. XIV in showing that inherent differences between male and female voices were not responsible for the effect shown in Fig. 14. Thus, in the present experiment, the stimulus lists were read by a female voice and the suffix presented either in the same voice or in a male voice.

Method.—The stimuli were lists of seven digits presented in a female voice. The Ss received two blocks of 29 digit lists in which the condition varied at random between a same (female) voice suffix and a different female-voice suffix and a male-voice suffix. The stimuli were played over a loudspeaker to 17 Ss. All the other procedural details were the same as in Exp. VIII–XII.

Results.—The pooled data are shown in Fig. 16. Wilcoxon tests showed that the effect of the same-voice suffix was different only at Position 7 from the other female voice ($p < .05$) and from the male-voice

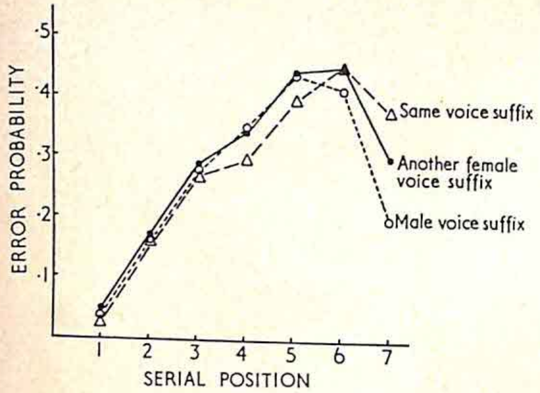

FIG. 16. Data from Exp. XV showing that with the stimuli spoken by a female voice, a suffix in the same voice causes more errors than a suffix in a different female voice. (A male-voice suffix has even less effect.)

suffix ($p < .01$). The (nonsignificant) advantage for the Same-Voice condition at Positions 4 and 5 is probably attributable to grouping effects and serves to emphasize the decrement at the final serial position. Thus, the previous result is confirmed and generalized to either relation between the male and female voices, and to differences between voices of the same timbre.

Experiment XVI

Given that the difference between male and female voices and differences among female voices lead to differences in the effect of the suffix, it is also of interest to see whether differential effects could be found when the pitch or accent of the suffix was varied.

Method.—In this experiment, there were six conditions each involving 18 seven-digit lists. The stimuli were recorded in a female voice. The six conditions are as follows: (*a*) Same-Voice suffix; (*b*) suffix in the same voice as the stimuli but at a different pitch, about a third higher; (*c*) and (*d*) the suffix was recorded in two other female voices, one American and one English; (*e*) and (*f*) two male voices were used for the suffix, American and English, respectively. There were three groups of four *S*s who heard the conditions in the orders: 531462, 162534, and 425613, respectively. Otherwise, the method was the same as in the preceding study.

Results.—The only significant differences among any of the conditions occurred in the final serial position. The same-voice suffix produced the greatest effect, being significantly different from the other two female voices ($p < .02$ and $p < .05$, two-tailed), from the same voice with different pitch ($p < .02$) and from the two male voices ($p < .01$ in both cases). This result confirms the previous ones and adds the fact that the pitch of the voice appears to be a feature which can be used by the selection mechanism. The lack of distinction between the accents might be because a single word "nought" carries insufficient acoustic cues to signal the difference.

Experiment XVII

Having shown that a suffix of a different pitch from the memory series has a reduced effect, we ought now to be able to reduce that advantage by presenting suffixes at different pitches at random in the same way that laterality effects were weakened by random conditions of presentation in Exp. X.

Method.—All details of method were the same as in the preceding study. Three levels of pitch were used, one the same as the stimulus, one a musical third higher, and a third pitched a third lower. There was a single block of 51 scored trials (seven-digit lists) preceded by practice items. These were played to a single group of 14 *S*s.

Results.—The only differences between the lists were between same and lower pitches on Position 5 ($p < .05$) and the lower pitch suffix caused fewer total errors than the upper pitch suffix ($p < .05$). These unpredictable differences can be attributed to chance. Thus, with a randomly varying pitch of the suffix, there were no differences germane to the suffix effect.

General Discussion

Let us now summarize our conclusions with respect to PAS. In the Crowder and Morton (1969) article, we claimed to have demonstrated the existence of an information store which gave a special advantage to auditory over visual presentation for the final items in a list. This advantage was removed by the presentation of a stimulus suffix, a redundant item which did not have to be responded to. The postulated information store, PAS, could be placed in and related to other information-processing func-

tions in the context of a more general model for information processing, the logogen model (Morton, 1969, 1970). It was predicted from this model that the intrinsic characteristics of the suffix should be irrelevant in determining the size of the suffix effect. Experiments II–V in the present paper confirmed this prediction with respect to meaning, frequency of occurrence, and emotionality. Crowder and Raeburn (1970) also showed that reversed speech yielded a suffix effect.

The later experiments reported above have investigated variables which do bear on the size of the suffix effect. Experiments VI and VII, together with one comparison from Exp. II, established that of the apparent spatial location of the suffix differs from that of the stimulus list, then its effect is reduced. The finding that a suffix presented to both ears is equivalent to a suffix presented to the opposite ear from the stimulus list established that the effect operates after the combination of information to the two ears. Experiment VIII showed that there were no artifacts in this result which could be ascribed to the loudness of the suffix with respect to the stimuli.

The importance of the voice characteristics of the suffix were shown in Exp. XII–XIV, in which it became apparent that the less the suffix resembled the stimuli, in timbre or pitch, the smaller the decrement in recall.

We have also claimed that a large part of the suffix effect can be attributed to events occurring after some attention mechanism. Our evidence for this is that the differences between suffixes can be reduced either by making their characteristics unpredictable (Exp. IX vs. VII for location and Exp. XIV vs. XV for pitch) or by forcing *S*s to process the suffix before recalling the digit string (Exp. XI vs. X). It might be noted here that the latter procedure did not succeed in producing an effect of a visual suffix on an acoustic stimulus or vice versa (Morton & Holloway, 1970). This result was also in accordance with the underlying model which asserts that prior to categorization, there is no overlap between the mechanisms involved in processing the two modalities.

We regard PAS as just one of a number of information stores which can be used in the course of an investigation into memory. It reveals itself maximally when *S*s are constrained to recall the stimuli in their order of presentation, where the contribution of PAS as shown both by the visual–auditory comparison and in the effects of the suffix are large and consistent. We have, however, also shown that it plays a role in the running memory span (Crowder & Morton, 1969, Exp. II) and in a design where *S* had to recall but one item, being given all the other items as a cue to recall. In these experiments, the effect of the suffix was numerically smaller than in the serial recall design (but no less significant) owing to the presence of other sources of information. These sources of information have been discussed at length by Crowder (1970) and Morton (1970).

It has recently become apparent that one change might be required in the model. We have claimed that PAS can only account for the advantage of auditory over visual presentation for the last few serial positions. However, Murdock and Walker (1969) have shown an advantage for auditory over visual presentation over the last five or six items of a free recall list. Since this difference existed for a rate of presentation of one English disyllable per 2 sec., Murdock and Walker argued that it is unreasonable to suppose it due to greater difficulties in processing the visual material. Murdock and Walker argued in favor of the existence of separate prelinguistic auditory and visual short-term stores with a persistence of up to 5 or 10 sec. If their arguments are accepted, then it is clear that such an acoustic store does not correspond to PAS since in the first place the auditory-visual difference is too small and second the difference covers too many serial positions and third there is evidence that PAS lasts no longer than 2 sec. (Crowder, 1969a, 1971). If it were shown in the free recall paradigm that a simulus suffix removed the advantage for acoustic presentation, then our opinion would have to be revised.

It might also be noted that PAS cannot account for the preperceptual auditory images described by Massaro (1970). Massaro showed that the identification of a pure tone was affected by a subsequent masking tone as a function of its closeness in time. However, unlike the suffix effect, this masking was unaffected by whether the masking tone was in the same or the opposite ear to the test tone.

In the course of the paper, we have noted that the suffix curves differ from control curves over the whole or most of the list. We have, however, referred to "the suffix effect" as restricted to the last few serial positions. Our justification for this procedure rests in the acknowledgment that the effect of a suffix is not restricted to "the suffix effect." When the suffix is presented, *S*s, perforce, process it and, at a subvocal level, respond to it. This implicit response should, according to the under-

lying model, act as a prefix. The effect of a prefix is to reduce performance in serial recall over the whole of the list (except the final item with auditory stimuli) as shown in Fig. 2. Crowder (1970) has shown that a comparison of suffix and prefix curves indicates that the suffix comparatively affects only the last two items.

We believe that PAS is now firmly established as a distinct theoretical construct. The effects associated with it can now be used as a tool to investigate phenomena such as attention and speech perception.

REFERENCES

BROADBENT, D. E. *Perception and communication.* New York: Pergamon, 1958.

BROADBENT, D. E. Word frequency effect and response bias. *Psychological Review*, 1967, **74**, 1–15.

BROADBENT, D. E., & GREGORY, M. Perception of emotionally toned words. *Nature*, 1967, **215**, 581–584.

BROWN, W. P. Conceptions of perceptual defense. *British Journal of Psychology*, No. 35, 1961.

CONRAD, R. Accuracy of recall using keyset and telephone dial, and the effects of a prefix digit. *Journal of Applied Psychology*, 1958, **42**, 285–288.

CONRAD, R. Very brief delay of immediate memory. *Quarterly Journal of Experimental Psychology*, 1960, **12**, 45–47.

CONRAD, R., & HULL, A. J. Input modality and the serial position curve in short-term memory. *Psychonomic Science*, 1968, **10**, 135–136.

CROWDER, R. G. Prefix effects in immediate memory. *Canadian Journal of Psychology*, 1967, **21**, 450–461.

CROWDER, R. G. Improved recall for digits with delayed recall cues. Paper presented at the meeting of the Psychonomics Society, St. Louis, November 1969. (a)

CROWDER. R. G. Phonic interference and the prefix effect. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 302–321. (b)

CROWDER, R. G. The role of one's own voice in immediate memory. *Journal of Cognitive Psychology*, 1970, **1**, 157–178.

CROWDER, R. G. Waiting for the stimulus suffix: Delay, decay rhythms and readout in immediate memory. *Quarterly Journal of Experimental Psychology*, 1971, in press.

CROWDER, R. G., & ERDMAN, H. P. Output interference in recalling digits. Paper presented at the meetings of the Eastern Psychological Association, Washington, D. C., April 1968.

CROWDER, R. G., & MORTON, J. Precategorical acoustic storage (PAS). *Perception & Psychophysics*, 1969, **5**, 365–373.

CROWDER, R. G., & RAEBURN, V. P. The suffix effect with reversed speech. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 342–345.

DALLETT, K. "Primary memory": The effects of redundancy upon digit repetition. *Psychonomic Science*, 1965, **3**, 237–238.

KIMURA, D. Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology*, 1961, **15**, 166–171.

KLEIN, G. S. Semantic power measured through the interference of words with color naming. *American Journal of Psychology*, 1964, **77**, 576–588.

MASSARO, D. W. Preperceptual auditory images. *Journal of Experimental Psychology*, 1970, **85**, 411–417.

MORTON, J. A preliminary function model for language behaviour. *International Audiology*, 1964, **3**, 216–225.

MORTON, J. Selective interference in immediate recall. *Psychonomic Science*, 1968, **12**, 75–76.

MORTON, J. The interaction of information in word recognition. *Psychological Review*, 1969, **76**, 165–178.

MORTON, J. A functional model for memory. In D. A. Norman (Ed.), *Models for human memory.* New York: Academic Press, 1970.

MORTON, J., & HOLLOWAY, C. M. Absence of cross-modal "suffix effects" in short-term memory. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 167–176.

MURDOCK, B. B., JR., & WALKER, K. D. Modality effects in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 665–676.

MURRAY, D. J. The effect of white noise on the recall of vocalized lists. *Canadian Journal of Psychology*, 1965, **19**, 333–345.

NATSOULAS, T. Converging operations for perceptual defense. *Psychological Bulletin*, 1965, **64**, 393–401.

ROUTH, D. A. "Trace strength," modality, and the serial position curve in immediate memory. *Psychonomic Science*, 1970, **18**, 355–357.

SHANKWEILER, D., & STUDDERT-KENNEDY, M. Identification of consonants and vowels presented to left and right ears. *Quarterly Journal of Experimental Psychology*, 1967, **19**, 59–63.

SHANKWEILER, D., & STUDDERT-KENNEDY, M. Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, 1970, **48**, 579–594.

TREISMAN, A. M. Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, 1960, **12**, 242–248.

TRIESMAN, A. M. The effect of irrelevant material on the efficiency of selective listening. *American Journal of Psychology*, 1964, **77**, 533–546. (a)

TREISMAN, A. M. Verbal cues, language and meaning in selective attention. *American Journal of Psychology*, 1964, **77**, 206–219. (b)

(Received May 26, 1971)

Journal of Experimental Psychology

Vol. 91, No. 2 — December 1971

TEMPORAL ASPECTS OF DIGIT AND LETTER INEQUALITY JUDGMENTS[1]

JOHN M. PARKMAN[2]

Carnegie-Mellon University

The time needed for adults to indicate which of two digits is larger, which of two digits is smaller, and which of two letters appears later in the alphabet was studied in three experiments. Latencies in both digit tasks were primarily a linear increasing function of the minimum digit of each pair, with responses for indicating the larger digit approximately 40 msec. faster than those for indicating the smaller digit. Latencies for the letter task were approximately 200–300 msec. longer than the reaction times for digits. Furthermore, the pattern of latencies for individual letter pairs was substantially different from the patterns for corresponding digit pairs, suggesting underlying process differences for the two types of material.

Moyer and Landauer (1967) investigated the time required to judge which of two simultaneously presented digits is larger and found latencies to be a monotonically decreasing function of the absolute numerical difference (Split) between the digits. Fairbank (1969) replicated this finding and, furthermore, found similar results when either months of the year or letters of the alphabet were substituted for digits. In addition, Fairbank showed that when *S*s are required to select the *smaller* of two simultaneously presented digits, more time is required than is needed to find the larger digit.

In interpreting their results, the authors of both articles suggested that the processes underlying inequality judgments between numbers are closely related to, and may be the same as, the processes involved in inequality judgments for physical continua. More specifically, they suggested that when *S* compares two numerals, he converts each numeral to an internal magnitude and a comparison is then made between these two magnitudes using an analog process in much the same way that *S* would make a comparison judgment between two physically continuous stimuli. The authors based this suggestion on the fact that such judgments for physical continua (e.g., colors differing in hue, lengths of lines, and frequency of tones) tend to be related to reaction time by functions which deviate from linearity in a manner similar to the deviation from linearity found in their own data.

Restle (1970) has found similar results

[1] This research was supported in part by United States Public Health Service, National Institute of Mental Health, Research Grant MH-07722. The author wishes to thank Patrick Cavanagh, William Chase, and Warren Throop for their critical comments. The assistance of Maria Gregory, Thomas Lee, and Sherry Sukol in data collection and analysis is gratefully acknowledged. The results of Exp. I and II were reported at the meeting of the Midwestern Psychological Association, Detroit, Michigan, May 1971.

[2] Requests for reprints should be sent to John M. Parkman, Department of Psychology, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213.

in a mental addition task. He investigated the time required to add two numbers $(p + q)$, compare the sum to a third number (r), and choose the larger. This task can be viewed as having two stages: a stage in which the sum is computed and a stage in which the result of this computation is compared to the third number. The task requirements of the comparison stage are very similar to those for the tasks studied in the articles mentioned above, although the range of numbers used in Restle's task was larger. (In general, two- or three-digit numbers were being compared.) Restle found that latencies monotonically decreased with increasing Splits between $(p + q)$ and (r) and adopted Moyer and Landauer's (1967) explanation of their data in interpreting his own comparison stage result. However, Restle also found that latencies decreased monotonically as the Split between p and q increased. This addition stage effect appeared strikingly similar to the comparison stage result. Thus, it is not surprising that Restle, in discussing possible processes underlying the addition stage, suggested that the magnitudes of p and q are converted to distances along a continuous internal number line in such a manner that the sum can be estimated from the length of the two connected line segments. One inference to be drawn from Restle's work, then, is that a strong possibility exists that the processes underlying mental addition are the same as, or are closely related to, the processes underlying inequality judgments between numbers and that these related processes are based on analog mechanisms.

However, Parkman and Groen (1971) have found evidence for digital-like processes underlying simple mental addition. In their task, *S*s added two single-digit numbers, compared the sum to a third number, and decided whether the third number and the sum were the same or different. Like Restle (1970), they found that reaction times are a monotonically decreasing function of the Split between addends, but they further found that over the range of problems they used, latencies are linearly related to the minimum of the two addends. In fact, the size of the minimum addend accounted for substantially more reaction time variance in the addition problems they studied than did the Split. This linearity in the smaller addend strongly suggests the presence of some type of counting mechanism embedded in the processes underlying the mental addition of at least single-digit numbers.

If the mechanisms underlying mental addition are, indeed, closely related to those underlying inequality judgments between numbers, then it should be possible to find linear effects in inequality judgment tasks similar to those found by Parkman and Groen (1971) for mental addition. The present series of three experiments was conducted to investigate this hypothesis.

The first experiment investigated the time required to choose the larger of two simultaneously presented digits and, hence, was a replication of the Moyer and Landauer (1967) study. However, in order to investigate this task for the presence of the type of linear effects found in simple mental addition, the data were analyzed not only as a function of the Split, as was done by Moyer and Landauer, but also as a function of other structural variables, viz., the minimum digit in each pair (Min), the sum of the two digits (Sum), and the maximum digit within each pair (Max). Of particular interest was whether reaction times would be a linear increasing function of the Min, as has been found for simple addition. A process model for the Moyer and Landauer digit-comparison task was constructed based on the results of this experiment and is presented in the Discussion.

The second experiment investigated the time required to choose the smaller of two simultaneously presented digits. As mentioned above, Fairbank (1969) found that the performance of this task requires more time than the performance of the comparable task using the "larger than" relation. Of particular interest in the present study was whether or not this increased time represents a change in process over the two tasks. One possibility is that performance using the "smaller than" relation is composed of two additive stages: one

stage in which the greater digit is detected, using the same process that is used in the task requiring selection of the larger digit, and a second stage in which the unselected (i.e., smaller) digit is found.

The third experiment investigated the relation of the Moyer and Landauer task to nonnumerical material. In this experiment, pairs of letters from the alphabet were presented simultaneously and *S*s were asked to find which of the two letters came later in the alphabet. Two sets of letters were used. One experimental group saw pairs of letters selected only from among the first 10 letters of the alphabet (A–J), and a second group saw letters selected only from the second 10 letters (K–T). The purpose of the experiment was to investigate the degree to which the processes underlying relational judgments of numbers are similar to or different from the processes used in making relational judgments of nonnumerical material.

Experiment I

The design used by Moyer and Landauer (1967) consisted of randomly presenting to *S* the 72 nonrepeating pairings of the digits 1–9, three times each. The design used for Group 1 of the present experiment differed only in minor details from this basic procedure. Specifically, *S*s were presented four blocks of trials, each block consisting of the 90 nonrepeating pairings of the digits 0–9, randomly ordered within certain sequential trial constraints.

The above designs are confounded in a number of ways. One source of confounding, noted by Moyer and Landauer (1967), is that the numerically larger digits have systematically higher probabilities of being correct, given the particular set of digit pairings they used. Furthermore, if these pairings are classified as to size of the Split, the Min, or the Max, it is found that these pairs tend to have systematically more small Splits than large Splits, systematically more small Mins than large Mins, and systematically more large maximum digits than small maximum digits, respectively. If any of these or related structural variables (e.g., the Sum) are important in determining reaction time, then this systematic bias in the frequency of occurrence of the respective values of such variables could, through repetition effects, confound the derived relations between these variables and latency of response.

To control for these possibilities, a second group (Group 2) was used in which the set of digit pairings presented differed from the set shown to Group 1. In particular, for Group 2 all possible nonrepeating pairings of the digits 0–9 were presented in which the digits of each pair differed from one another by not more than three, and all such pairings were presented equally often. This design generates a generally flat frequency distribution of digit pairings with regard to the Split, the Min, and the Max of individual pairs. In addition, the digits 3–6 occur equally often as the Min and as the Max within individual pairs, whereas the digits 0–2 systematically tend to be paired with larger digits and the digits 7–9 similarly tend to be paired with smaller digits. Hence, a check on the possibility of artifact resulting from the confounding of digit size with response is possible.

Method

Subjects.—The *S*s were 41 students from introductory psychology classes at Carnegie-Mellon University, who received class credit for participation. The *S*s were randomly assigned to the two experimental groups, resulting in 22 *S*s in Group 1 and 19 *S*s in Group 2.

Materials and apparatus.—The materials consisted of single digits ranging in value from 0–9. A single pair of such digits was displayed simultaneously on each trial with the constraint that the two digits could not be identical. The digit pairs were presented on a standard video monitor controlled by a DDP-116 digital computer. Each pair of digits appeared centered horizontally on the video screen with individual digits subtending .5° of visual angle and with center to center spacing between digits of 2.0°. The *S*s responded by pressing one of two microswitches, appropriately labeled "LEFT" and "RIGHT," mounted approximately 50 cm. in front of the video monitor. Latencies were measured by the computer to the nearest .001 sec. from the time the digits appeared on the screen until switch contact was made.

Procedure and design.—For both Group 1 and Group 2, individual trials were conducted as follows: (*a*) The message "READY" appeared centered on the screen for .5 sec. (*b*) The screen remained blank for .5 sec. (*c*) The two digits were displayed on the monitor

TABLE 1
INTERCORRELATION MATRIX FOR STRUCTURAL VARIABLES

Variable	Group 1			Group 2		
	Split	Min	Max	Split	Min	Max
Split						
Min	−.500			−.172		
Max	.500	.500		.172	.941	
Sum	.000	.866	.866	.000	.985	.985

until S made a response. (*d*) The screen remained blank for 1.0 sec. prior to the warning for the next trial. In the case that S responded incorrectly, the above procedure was modified in the following manner: (*i*) The screen remained blank for .5 sec. (*ii*) The message "WRONG" appeared centered on the monitor screen for 1.0 sec. (*iii*) The screen remained blank for 1.0 sec. prior to the warning for the next trial.

The two groups differed in their respective set of digit pairs used as stimuli. For Group 1, all possible nonrepeating pairings of the digits appeared. The computer randomly permuted these 90 possible pairings within blocks for each S with the constraint that no single digit could appear on successive trials. Four such blocks were presented to each S with an interblock rest interval of 20 sec. (During rest periods, the message "REST" appeared centered on the monitor screen.) For Group 2, all possible nonrepeating pairs of digits were used in which individual digits differed from one another by not more than three. The computer randomly presented these 48 possible pairings within blocks for each S, again using the constraint that no single digit could appear on successive trials. Seven such blocks were presented to each S, with interblock rest intervals identical to those used for Group 1.

The Ss served in one session each. The instructions to both groups asked S to decide on each trial which of the two digits was numerically larger and to press the right switch if the larger digit occurred on the right side, and vice versa. In addition, all Ss were asked to respond "as quickly and accurately as possible" and to use their index fingers for responding.

Results

Structural variables.—In analyzing reaction times, only data obtained from trials on which a correct response was made were considered. The mean reaction times for the individual digit pairs for each of the two experimental groups were investigated with respect to four structural variables—the Split, the Min, the Sum, and the Max. The goal of this analysis was to find which one or more of these variables is most closely related to reaction time in the present task with the belief that such information would provide clues toward the identification of the processes underlying inequality judgments between numbers.

In performing such an analysis, a problem arises with regard to the a priori intercorrelations among the structural variables studied. Assume that one set of structural variables is strongly related to reaction time and that a second set of such variables is not. If any of the second set of variables is highly correlated with any of the variables in the first set, then the members of the second set will appear to be related to reaction time also. For example, assume that reaction times in the present task are a function of the Min only. If a design is used in which the Split and Min are correlated, the Split will be spuriously correlated with reaction time to the approximate extent that it is correlated with the Min. Such a situation would hinder attempts to identify the mental processes underlying the present task. However, if two or more experimental designs are used and the intercorrelations among the structural variables of interest differ across these designs, a resolution of this problem is sometimes possible. In general, the correlations with reaction time for the nonspuriously correlated variables will not significantly change across the various designs while the same correlations for the other variables will rise and fall in magnitude in direct correspondence to the pattern of correlations of these variables with the more stable variables. Although this is precisely true only in an ideal case, if the number of structural variables studied is sufficiently small, this argument can be used as a heuristic for identifying the more important variables for a given task, given that the variables investigated are not independent.

Table 1 presents the a priori correlations among the four structural variables investigated based on the designs used for Group 1 and Group 2. Table 2 shows for both groups the intercept, slope, and R^2 values for each of these structural variables using the method of least squares regression. It can be seen from Table 2 that the Min

accounts for approximately 77% of the variance in digit-pair means for both Group 1 and Group 2. Using the reasoning presented above, it can be argued that the relation of the Max and the Sum to mean reaction time stems from these variables' rather high a priori correlation with the Min. Similar argument can be applied to the Split. However, since the Split has been previously implicated as an important variable in the present task, closer examination is warranted. Using the method of semipartial correlation (Nunnally, 1967), estimates can be made of the relation between the Split and mean reaction time in the two designs after the Min has been partialed from the Split. When this analysis is performed, using the data in Tables 1 and 2, it is found that the Split accounts for approximately 7% of mean reaction time variance in Group 1 and 9% of the variance in Group 2. On the other hand, using the same technique, the Min accounts for an average of 72% of the variance in these two groups if the Split is partialed from the Min. (When similar estimates based on this procedure are made for the Sum and the Max, neither of these structural variables, in either experimental group, accounts for more than 9% of mean reaction time variance independently of the Min.)

TABLE 2

INTERCEPT (MSEC.), SLOPE (MSEC/UNIT), AND R^2 VALUES FOR STRUCTURAL VARIABLES

Variable	Group 1			Group 2		
	Intercept	Slope	R^2	Intercept	Slope	R^2
Split	481.5	−9.8**	.437	491.5	−15.5**	.198
Min	411.3	12.9**	.762	424.3	10.6**	.776
Max	425.9	3.1*	.045	414.1	8.7**	.530
Sum	397.7	5.3**	.392	417.0	5.0**	.667

* $p < .05$.
** $p < .01$.

It is clear then that the Min accounts for a substantial amount of reaction time variance in this task and that previous studies have overlooked its importance. In Fig. 1, mean digit-pair latencies and error rates are presented as a function of the Min. (Each point represents the pooled mean of all digit pairs with the appropriate Min value.) There is a high degree of linearity in these data. Linear regression (unweighted) accounts for 95% of the Group 1 reaction time variance and 96% of the variance for Group 2. The equations of the best-fitting (unweighted) straight lines describing these two set of data are given in the figure.

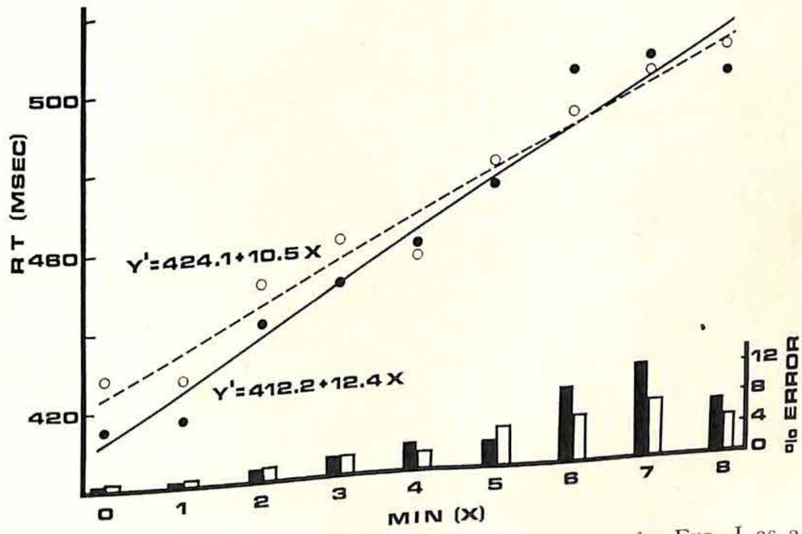

FIG. 1. Mean digit-pair reaction times and error rates for Exp. I as a function of the Min. (Filled circles and bars refer to Group 1, open circles and bars refer to Group 2.)

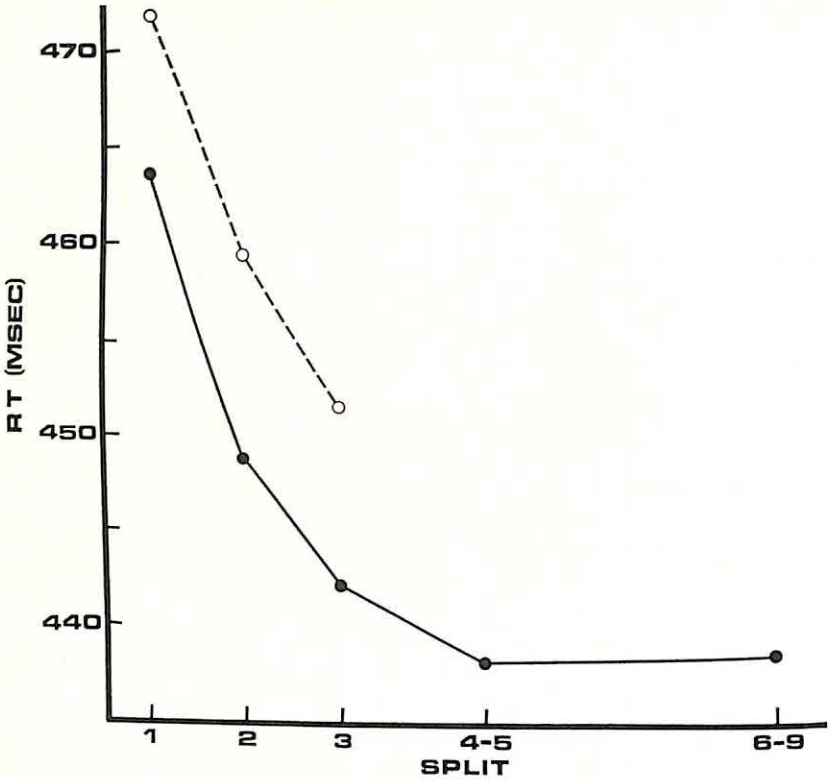

FIG. 2. Mean adjusted Min residual reaction times for Exp. I as a function of the Split. (Filled circles and bars refer to Group 1, open circles and bars refer to Group 2.)

Residuals.—Possible processes related to the Min will be presented in the Discussion. However, it is important to note that although the Min accounts for a substantial amount of reaction time variance in the present task, other structural variables may still account for an important variance component that is independent of the Min. One way to investigate this possibility is to examine the residual reaction times for the digit pairs when the Min has been used as the independent variable in regression (Draper & Smith, 1966). If any of the other structural variables are having systematic effects on reaction time independently of the Min, these effects should appear as systematic trends in these residuals. When the Min residuals were plotted as a function of the Max, the Sum, and the Split, no systematic trends could be found as a function of either the Max or the Sum. However, a clear, nonlinear, trend was discovered as a function of the Split. In particular, it was found that for Group 1, digit pairs with Split values of 1, 2, or 3 had systematically higher residual reaction times than did digit pairs with Split values of 4–9, which showed no systematic effects. Group 2 (which had Split values of only 1, 2, and 3) showed a similar pattern in its Min residuals to that found for Group 1. Figure 2 presents these Min residuals for both groups. (Note that these "residuals" have been adjusted by adding to each residual the mean reaction time for its respective experimental group.)

In Fig. 3, mean digit-pair latencies and error rates are presented as a function of the Split. This figure appears very similar to the corresponding figures presented by Moyer and Landauer (1967) and Fairbank (1969). Hence, the number inequality judgment effect, first investigated by Moyer and Landauer, is replicable. However, as shown above, the curves in Fig. 3 can be thought of as composed of two components: (*a*) a strong, linear component based on the Min and (*b*) a smaller, nonlinear component

based on that portion of the Split which is independent of the Min.

Order effects.—The individual digit-pair means for both groups were investigated for possible order effects with regard to the respective left–right positions of the Max and the Min of each pair. For both groups, it was found that this mean order difference amounted to less than 3 msec., with the direction of the "effect" being opposite in the two groups. These mean differences were not significant for either Group 1, $t\ (88) = .29,\ p > .50$, or for Group 2, $t\ (46) = .34,\ p > .50$.

Errors.—The correlation between errors and reaction time, based on individual digit-pair means, was .775 for Group 1 and .845 for Group 2. Although these correlations are relatively high, the actual error rates for both groups were quite low, being 2.86% for Group 1 and 1.92% for Group 2.

As mentioned above, the designs used for Group 1 and Group 2 substantially differed. Nevertheless, as can be seen in Fig. 1, 2, and 3, both these designs yielded very similar data. Hence, the possibility is greatly reduced that the design used for Group 1 (and employed by Moyer and Landauer) yields artifactual data due to the confounding mentioned earlier. Because of this finding, a decision was made to employ the general design used for Group 1 of the present experiment in the experiments reported next.

Experiment II

One way to explore the generality of the results of Exp. I is to present *S*s with the same material, but ask them to perform a somewhat different operation. In the present experiment, such an attempt was undertaken by asking *S*s to find the smaller of two simultaneously presented digits. As mentioned above, Fairbank (1969) has found that more time is required to find the smaller of two single digits (1–9) than to find the larger digit. He asked *S*s to scan double columns of digits and to mark through either the larger or the smaller digit in each row. Using a within-*S*s design, he found that an average of 24 msec. additional time per problem was required to find the smaller digit than was needed to find the larger digit. However, Fairbank's procedure did not provide information as to the relation of reaction time in this task to the structural variables investigated in Exp. I. Such information is needed in order to investigate possible process differences between the two tasks.

One theory that is relevant to possible process differences between finding the larger and finding the smaller of two digits is the "Principle of Lexical Marking," which

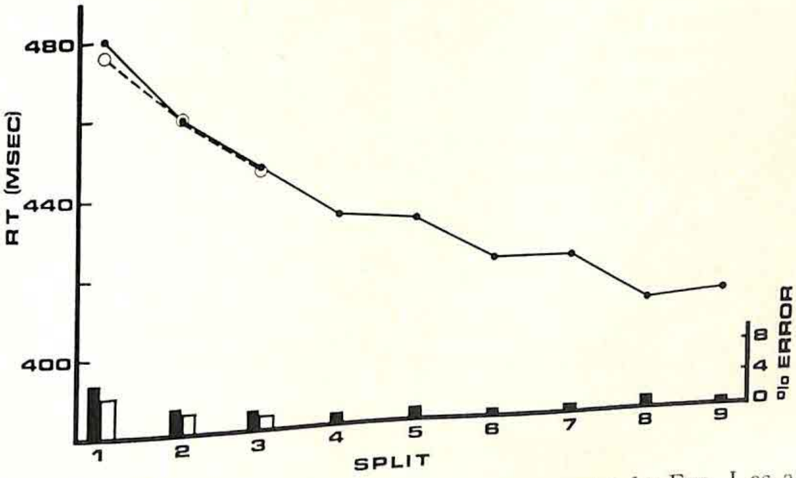

FIG. 3. Mean digit-pair reaction times and error rates for Exp. I as a function of the Split. (Filled circles and bars refer to Group 1, open circles and bars refer to Group 2.)

TABLE 3

INTERCEPT (MSEC.), SLOPE (MSEC/UNIT), AND R^2 VALUES FOR STRUCTURAL VARIABLES

Variable	All digit pairs			Digit pairs containing no zeroes		
	Intercept	Slope	R^2	Intercept	Slope	R^2
Split	523.5	−10.2*	.338	525.8	−11.6*	.339
Min	454.1	11.9*	.678	428.1	17.7*	.789
Max	475.2	1.7	.009	446.4	6.1*	.094
Sum	445.0	4.5*	.200	407.7	7.9*	.476

* $p < .01$.

has been presented by Clark (1969). According to this principle, antonymous adjectives, like large and small, are asymmetric in the sense that the positive or "unmarked" adjective ("large" in this example) is stored in memory and handled in a less complex way than is its opposite. In the present case, "smaller" should be coded in a manner analogous to "larger" but with the addition of one semantic feature or marking. The process effect of this marking for the present task might be to initiate retrieval of the unselected (i.e., smaller) digit after the process which finds the larger digit has made its selection. If this reasoning is correct, and if it is assumed that the time required for this extra operation is a constant regardless of the digits involved, then reaction times associated with finding the smaller of two digits should be the same as those associated with finding the larger digit, except for a small increment. Hence, reaction times for finding the smaller digit should be related to the structural variables of Exp. I in the same way as were the reaction times for finding the larger digit, except for small intercept differences. These possibilities were investigated in the present experiment.

Method

Subjects.—The *S*s were 23 students drawn from the same *S* pool as in Exp. I. The *S*s were given class credit for participation. No *S*s served in both Exp. I and II.

Materials, apparatus, procedure, and design.—The conduct of this experiment was identical to that for Group 1 of Exp. I except that in this experiment *S*s were asked to find the numerically "smaller" of the two digits and to press the right switch if the smaller digit were on the right, and vice versa.

Results

The data from this experiment were submitted to essentially the same analysis as were the data from Exp. I. Since the design used for the present experiment was basically the same as that used for Group 1 of Exp. I, a comparison of the results from these two sources should yield information as to the similarities or differences between the processes used by *S*s in the two tasks. As is shown below, the data from the present experiment do appear highly similar to the data from Exp. I, with one notable exception. Digit pairs which contained a zero had higher mean latencies in the present experiment than would have been predicted from the results of Exp. I. This was particularly true of the Pairs 0–1 and 1–0. Speculations relevant to this result are presented in the Discussion. However, in what immediately follows, an attempt is made to present the results in such a manner that any possible confounding with a "zero effect" is avoided.

Structural variables.—Table 3 presents the intercept, slope, and R^2 values, based on the individual digit-pair means of the present experiment, for each of the structural variables studied in Exp. I. Two separate sets of analyses have been performed for Table 3. One set of analyses used all 90 of the digit-pair means, while a second set used only the 72 digit-pair means based on pairs which did not contain a zero. With regard to the former analysis, it can be seen that the pattern of relationships that data from the present experiment have with the four structural variables is very similar to the pattern found for Group 1 of Exp. I, except that each of these variables accounts for roughly 10% less variance in the present experiment. Of particular importance are the following findings: (*a*) the Min accounts for a substantial portion of the mean digit-pair variance, as it did in Exp. I, and (*b*) the slope values for the Min in the two experiments are very similar.

Turning now to the second analysis, two important effects of removing digit pairs containing a zero can be seen. First, the Min accounts for 78.9% of the variance for

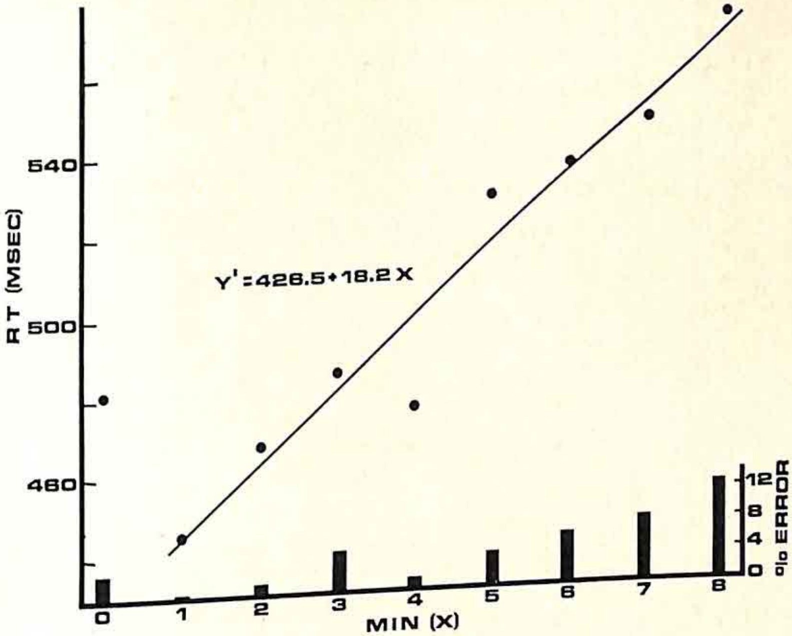

FIG. 4. Mean digit-pair reaction times and error rates for Exp. II as a function of the Min. (The data point at the zero Min value was not used in computing the regression line.)

the remaining digit-pair means, which is very similar to the corresponding figures for both groups of Exp. I. (The R^2 values for Group 1 and Group 2 of Exp. I, with digit pairs containing a zero removed, are .743 and .761, respectively.) Second, the slope for the Min increases to 18.2 msec/unit. (The corresponding slope values for Group 1 and Group 2 of Exp. I, with pairs containing a zero removed, are 13.7 and 11.2 msec/unit, respectively.)

In Fig. 4, mean digit-pair reaction times are presented as a function of the Min. The regression line (unweighted) in this figure has been fitted to all points, except for the point corresponding to a Min of zero. This line accounts for 95% of the variance in the nonzero points, indicating a high degree of linearity in the data when plotted in this fashion.

Residuals.—An analysis of the Min residuals for the digit pairs not containing a zero was performed for the present experiment, just as was done for Exp. I. No systematic trends in these residuals could be found as a function of either the Max or the Sum. However, these Min residuals plotted as a function of the Split revealed a clear, nonlinear trend very similar to that found in the first experiment. Figure 5 presents the Min residuals (adjusted as in Exp. I) from the present experiment as a function of the Split.

Larger versus smaller.—As can be seen above, the data from this experiment are quite similar to the data from Exp. I in terms of the structural variables investigated. In addition, further analysis showed that *S*s in the present experiment required significantly more time to indicate which was the smaller digit in each pair than was required by *S*s in Exp. I to indicate the larger digit. The *S*s of Group 1 of Exp. I had a mean latency of 444.3 msec., while *S*s of the present experiment averaged 484.3 msec. The 40.0-msec. difference between these groups was significant, t (43) = 3.09, $p < .005$. The standard deviations of the mean reaction times for *S*s of the two groups were 36.0 and 49.4 msec. for Group 1 of Exp. I and for the present experiment, respectively. Cochran's test for heterogeneity of variance did not detect a significant difference between the variances associated with these two figures, C (22) = .52, $p > .05$. The

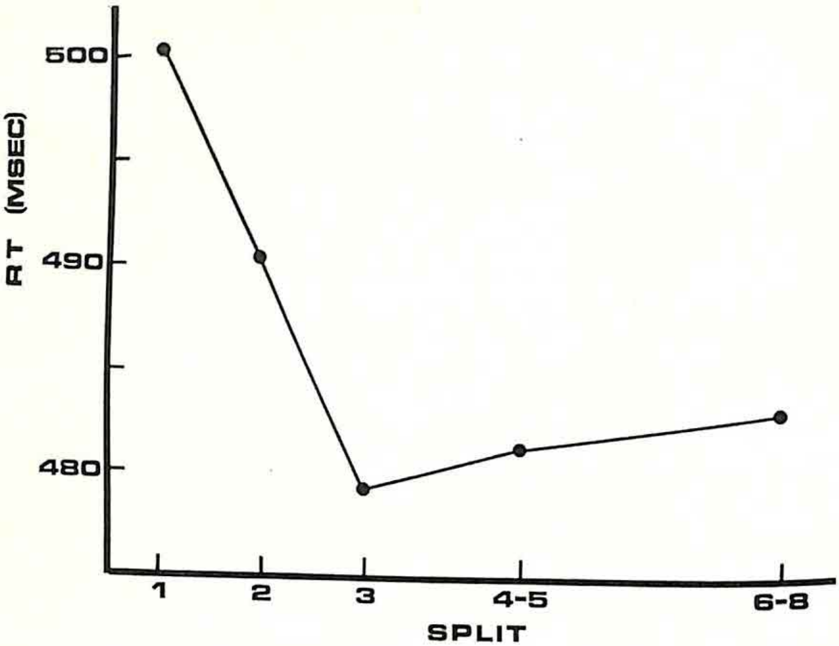

FIG. 5. Mean adjusted Min residual reaction times for Exp. II as a function of the Split.

reaction time differences between these two groups can also be viewed in terms of overall differences based on individual digit-pair means. Using all 90 digit pairs, the overall mean latency for Group 1 of Exp. I was 445.7 msec., while the corresponding mean for the present experiment was 485.9 msec. This 40.2-msec. increase in mean latency for the present experiment was significant, $t\ (178) = 7.46$, $p < .001$. If digit pairs with a zero are removed, the overall means of Group 1 of Exp. I and for the present experiment are 453.3 and 487.2 msec., respectively. This 33.9-msec. difference is also significant, $t\ (142) = 5.68$, $p < .001$. Although the possibility of *S* sampling bias cannot be ruled out, these estimates of the mean latency difference in the two tasks are of the same direction and order of magnitude as that found by Fairbank (1969), whose data, based on a within-*Ss* design, were less susceptible to possible sampling bias than were the data from the present study.

Order effects.—An investigation for possible Min and Max order effects in the digit-pair means for the present experiment was conducted. Although pairs with the Max in the left position had, on the average, 6 msec. longer mean latencies than was the case for pairs with the Max in the right position, this difference was not significant, $t\ (88) = .746$, $p > .50$. Removal of pairs containing a zero did not change this finding.

Errors.—The correlation between errors and reaction time in the present experiment, based on individual digit-pair means, was .779 for the 90 digit pairs. The mean error rate for the experiment was 2.81%.

EXPERIMENT III

This experiment investigated the extent to which the results of Exp. I and II can be extended to similar tasks using different materials. If the processes underlying the results of Exp. I are used only with numbers, then there is no reason to expect similar results from performance on an identical task with nonnumerical material.

The present experiment was identical to Exp. I except that alphabetically contiguous letters were substituted for the 10 digits and *Ss* were asked to indicate for each letter pair which letter appeared later in the alphabet. The results from the experiment were analyzed in terms of the same structural variables used in Exp. I and II by substituting zero for the alphabetically first letter, one for the next letter, etc. Emphasis was placed on determining to what extent the pattern of relations found

between letters and these structural variables are similar to the relations found for digits.

Using a design very similar to that used by Moyer and Landauer (1967) with digits, Fairbank (1969) presented nonrepeating pairs of letters drawn from the first nine letters of the alphabet (A–I) and asked *S*s to choose the letter in each pair which came later in the alphabet (i.e., nearer Z). He found reaction times to such letter pairs to be significantly related to the Split. However, in a second experiment using the same letters, Fairbank found that *S*s could find the letter nearer A faster than they could find the letter nearer Z. He concluded that while there may be similarities between finding the larger of two digits and finding the alphabetically later of two letters, the underlying decision processes are not identical.

One reason Fairbank (1969) found similarities in his results between digits and letters may have been due to the fact that he used the initial letters of the alphabet. Fairbank's *S*s (college students) probably knew the serial positions of these initial letters fairly well and could perhaps perform the task using this associated numeriical information. One hypothesis then is that Fairbank would not have found as close a similarity in his results between digits and letters if he had used letters drawn from the middle of the alphabet where presumably the letters' serial positions are not as well known. The present experiment used two groups of *S*s. Group 1 was presented nonrepeating letter pairs drawn from the first 10 letters of the alphabet (A–J), and Group 2 was presented letters drawn from the second 10 letters of the alphabet (K–T).

Method

Subjects.—The *S*s were 32 students drawn from the same *S* pool as in Exp. I. The *S*s were randomly assigned to the two experimental groups, resulting in 15 *S*s in Group 1 and 17 *S*s in Group 2. Again, *S*s were given class credit for participation, and no *S* who served in the earlier experiments served in Exp. III.

Materials.—The materials for Group 1 consisted of the first 10 letters of the alphabet, A–J. For Group 2, the 10 consecutive letters K–T were used. The letters were displayed in uppercase, block form and subtended a visual angle of .5° with a center-to-center spacing between letter pairs of 2.0°.

Apparatus, procedure, and design.—The apparatus, procedure, and design used in this experiment were identical to that for Group 1 of Exp. I, except that in this experiment *S*s were asked to find the letter in each pair coming later in the alphabet and to press the right switch if the later letter appeared on the right, and vice versa. They were also told before the experiment the range of letters that would be used in forming the pairs.

Results

The results of the present experiment with letters were found to differ in a number of important ways from the results of Exp. I and II, which used digits. One important difference was a substantial increase in overall mean latencies for the letters. These mean reaction times for the present experiment, based on individual *S* means, were 691.8 msec. for Group 1 and 743.5 msec. for Group 2. These means constituted an increase of more than 200 msec. in mean latency for the tasks using letters as opposed to the equivalent tasks using digits. The 51.7-msec. mean latency difference between Group 1 and Group 2 of the present experiment was significant, $t\ (30) = 1.78$, $p < .05$, one-tailed test.

Structural variables.—As a means for further comparison between letters and digits, Table 4 presents the intercept, slope, and R^2 values, based on the individual letter-pair means of the present experiment, for each of the structural variables studied in the previous experiments. It can be seen that the Min (i.e., the letter in each pair occurring earlier in the alphabet) no longer accounts for as substantial an amount of mean reaction time variance as was true for digits. Indeed, for Group 2, the Min accounts for almost no variance at all. The Split now

TABLE 4

INTERCEPT (MSEC.), SLOPE (MSEC/UNIT), AND R^2 VALUES FOR STRUCTURAL VARIABLES

Variable	Group 1			Group 2		
	Intercept	Slope	R^2	Intercept	Slope	R^2
Split	767.4	−25.6*	.493	864.2	−29.7*	.446
Min	616.4	21.4*	.346	776.4	−7.9	.032
Max	699.8	−4.1	.013	993.7	−37.7*	.716
Sum	621.7	5.8*	.075	892.0	−15.2*	.350

* $p < .01$.

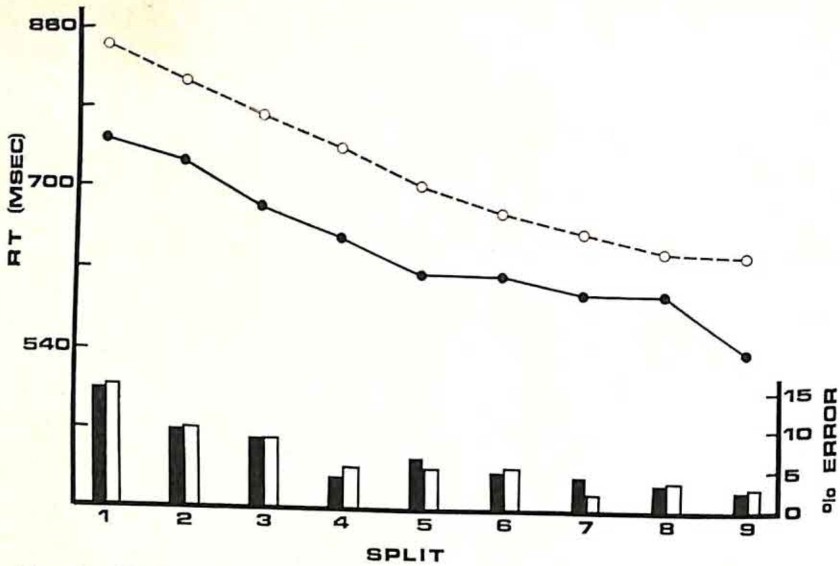

FIG. 6. Mean letter-pair reaction times and error rates for Exp. III as a function of the Split. (Filled circles and bars refer to Group 1, open circles and bars refer to Group 2.)

accounts for approximately half of the variance in the two groups. The Max accounts for a substantial amount of variance for the second 10 letters of the alphabet (Group 2) but accounts for very little variance in the first 10 letters (Group 1). The patterns of relations that Group 1 and Group 2 of the present experiment have with the four structural variables are very interesting in themselves but are clearly different from the corresponding patterns seen for digits in the first two experiments.

Figure 6 presents the mean letter-pair reaction times and error rates for Group 1 and 2 of the present experiment as a function of the Split. Of interest is the result that these two curves are very similar, except for intercept differences. Figure 7 presents the mean letter-pair reaction times and error rates for Group 2 as a function of the Max. Linear regression accounts for 97% of the variance in this figure.

Order effects.—An investigation for possible Min and Max order effects in the letter-pair means was conducted for the present experiment. With the Min in the left position (i.e., when the letters were in alphabetical order), reaction times were faster for both groups, on the average, than when the Min was in the right position. The size of this result was 19.7 msec. for Group 1 and 20.9 msec. for Group 2. However, neither of these results reached statistical significance, $t\ (88) = 1.15$, $p > .30$ for Group 1 and $t\ (88) = 1.00$, $p > .40$ for Group 2.

Errors.—The correlation between errors and reaction time in the present experiment, based on individual letter-pair means, was .657 for Group 1 and .654 for Group 2. The mean error rates were 8.27 and 8.32% for Group 1 and Group 2, respectively.

DISCUSSION

The most important results of this study with respect to digits are the following: (*a*) Reaction times for finding and indicating either the larger of two simultaneously presented digits or for finding and indicating the smaller digit are to a large extent a linear function of the Min, with the Min accounting for a substantial amount of individual digit-pair reaction time variance in both tasks. (*b*) For both of these tasks, those portions of digit-pair reaction times which are not explained by the Min are a monotonically decreasing function of the Split, with this "Split effect" being greater for digit pairs with relatively small Splits. (*c*) The task of finding the smaller of two simultaneously presented digits yields results very similar to the task of finding the larger digit, the only major difference being a constant small increase in reaction time.

With regard to letters, the most important results are the findings that (*a*) reaction times for indicating which of two simultaneously presented letters appears later in the alphabet are more than 200 msec. longer than was found

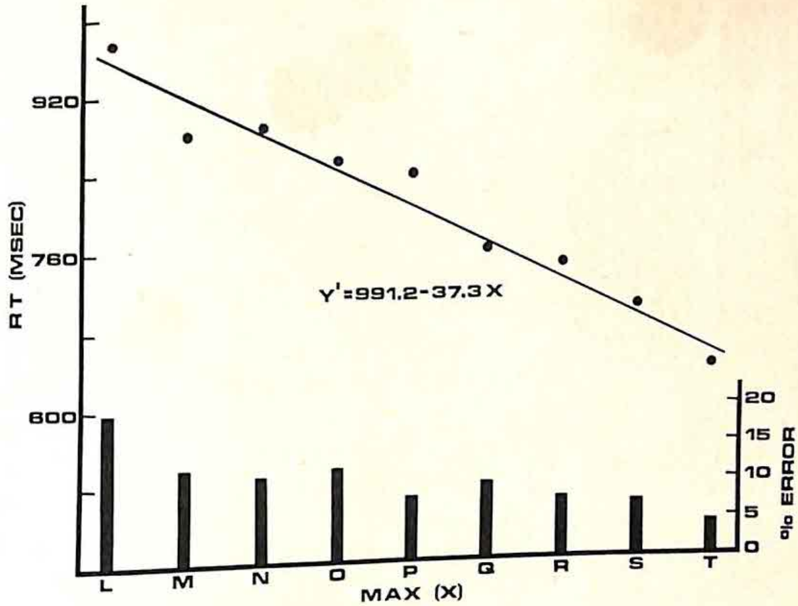

FIG. 7. Mean letter-pair reaction times and error rates for Group 2 of Exp. III as a function of the Max. (The regression equation assumes that the numbers 1–9 have been substituted for the letters L–T, respectively.)

for the corresponding task of selecting the larger of two digits, and (*b*) the pattern of reaction times for letters substantially differs from the pattern found for digits.

Attempts to explain the above results are frustrated by the rather primitive level of current theories of human information processing. In fact, at the level of analysis of the present study, no theory today proposes how such an elementary task as that of finding the larger of two digits is performed. In what follows, therefore, no attempt is made to examine the results of the present study in the light of existing theory. Rather, process models based on these results are derived and presented. The psychological significance of the components of process models such as these have yet to be determined, but will, hopefully, be the subject of theory in the near future.

Any model of the digit inequality judgment tasks investigated in this study should accomodate the above findings in regard to the Split and, especially, the Min. One type of model which is compatible with the Min finding is a counting model built around the notion of a counter or incrementing register. One such counting model which can accomodate the above Min result is the following, which is tentatively being proposed for the inequality judgment task of Exp. I. (An extension of this model will later be proposed to encompass the results of Exp. II.) First *S* encodes the two numerals. Next he activates an internal match register in the following manner. The register is initially set to zero. After encoding the digits, the register is incremented by one, and the contents of the register are compared against the representations of the two encoded digits. (As there are no order effects in the data, this match against the two representations may occur either serially and exhaustively or by means of some parallel comparison scheme.) If a match occurs, the increment and compare procedure is terminated, and the *unmatched* representation is retrieved. In case a match does not occur, the match register is incremented by one and another comparison is made. This increment and compare procedure continues until a match is found. Finally, it is assumed that retrieval of the selected digit provides access to information concerning that digit's relative position in the original visual field in order that the proper motor response may be initiated.

The number of increment and compare cycles required by this model for any particular digit pair is, of course, equal to the Min of that pair. Hence, the Min slope is an estimate of the time required for one complete operation of this testing procedure. The 10–20 msec/unit Min slopes found in Exp. I and II mean that between 50 and 100 increment and compare cycles are made in a single second.

The effect of the Split on digit-pair latencies can be incorprated into the present model in the following, if somewhat post hoc, fashion.

The comparison procedure mentioned above may well operate in a signal detection manner. Without going into the details of the Theory of Signal Detection (Green & Swets, 1966), it can be argued that for the present case, when two-digit encodings are sufficiently similar (i.e., when the Split between two digits is sufficiently small), the execution of the above comparison procedure is made more difficult. This added difficulty would be reflected in increased latencies and errors for digit pairs with small Splits. Results supporting this argument can be seen in Exp. I and II, although further confirming evidence is obviously needed before this "comparison procedure effect" can be entertained as more than a speculation.

The results of Exp. II also are compatible with the model presented above if it is assumed that to choose the smaller digit, *S* must first find the larger digit and then select the smaller digit by retrieving the digit that was not selected by the process which found the larger digit. Using the results of Exp. I and II, the time required for this extra operation in retrieving the smaller digit can roughly be estimated as approximately 40 msec. (Actually, one should compare intercept differences in making this estimate as the model predicts that the Min slopes for both tasks will be the same and that the result of the additional mental operation needed in Exp. II will be just an increase in the intercept for the data from that experiment when plotted as a function of the Min. However, since means are more stable than intercepts, the means from Group 1 of Exp. I and from Exp. II have been compared for making this estimate.)

As the model predicts that *S* actually finds the smaller digit (the Min) first in order to identify the larger digit, it would appear inefficient to find the smaller digit by first finding the larger. One possibility, though highly speculative, is that the process of finding the larger digit (the normal mode of operation for the model) somehow temporarily leaves the representation of the smaller digit in some kind of refractory state. Hence, the smaller digit cannot be identified directly but only in reference to the larger digit.

With regard to digit pairs containing a zero, the model generates the *wrong answer*. This stems from the fact that the match register can never detect a zero, as it increments before it tests for a match. Hence, the "effective Min" (as far as the model is concerned) for pairs with zeroes is, in reality, the Max. Many *S*s in Exp. I and especially Exp. II reported that they had difficulty with zeroes. It may be that *S*s developed strategies for handling digit pairs containing zeroes which were unique to this kind of pair and which, for Exp. II, demanded more time than did the processing of pairs without zeroes. However, the nature of such strategies are not at present understood. Nor is it known why this "zero effect" should be stronger in Exp. II than in Exp. I. One hint may be that zeroes were never "correct" in Exp. I, but what this means in terms of process is not at present clear.

Whether or not the above model is an accurate description of the processes underlying the selection of minimum and maximum digits, the high degree of similarity found in the results of Exp. I and II certainly lends strong support to the argument that whatever the processes are that underlie these two operations, they are very similar. Nevertheless, a within-*S*s investigation of these two tasks, using procedures similar to those used for Exp. I and II still needs to be conducted to check for possible Min slope differences and to provide a better estimate of overall mean response latency differences.

Turning now to letters, the results of Exp. III clearly indicate that there are significant differences between finding which of two digits is larger and which of two letters comes later in the alphabet. The most important of these differences in the data derived from these two tasks are the increase in overall mean latencies for the letter task and the fact that mean letter-pair latencies are not systematically related to the Min as is the case with digits. Still, the results of Exp. III are interesting in their own right. Although mean letter-pair latencies are a decreasing function of the Split both for pairs composed of letters which appear early in the alphabet (A–J) and for pairs composed of letters appearing later (K–T), there are also marked position effects in the results. More specifically, there is a fairly strong Min effect for Group 1 while there is a very strong Max effect for Group 2. One interpretation of these effects along with the finding in regard to the Split is that *S*s perform these letter tasks by keeping a short list of contiguous letters from the alphabet in short-term memory and checking on each trial to see if one of the presented pair of letters is a member of this list. Presumably this "short list" is composed of initial letters of the alphabet for Group 1 and letters near T for Group 2. If one of the two letters is a member of such a set, then *S* will have information with which to solve the problem of which letter comes later in the alphabet. However, if neither or both of the displayed letters are part of *S*'s "memory set," then the list has to be expanded or shortened until only

one of the displayed letters is contained within the list. This post hoc "model" predicts that mean letter-pair latencies will have the general properties found in Exp. III, but further work related to this experiment clearly needs to be done.

In a recent article, Lovelace and Snodgrass (1971) reported the results of a study whose purpose was to extend the work done by Moyer and Landauer (1967) "to decisions involving alphabetic order of letter pairs presented simultaneously [p. 258]." In their tasks, *S*s had to indicate whether a pair of letters was in proper or reversed alphabetic order. Among other results, they found that latencies decreased as the Split between letters increased and that latencies were shorter for letter pairs in proper than reversed alphabetic order. Lovelace and Snodgrass concluded that their results extended "the generality of Moyer and Landauer's (1967) finding with digits to the case of the order property of letters [p. 262]." However, closer examination of their results, especially in view of the results of the present study, complicates this conclusion. While their letter-pair means do generally decrease with increasing Splits, it is not clear what the effect of the Min of each pair is on their data. Also, there are significant order effects in their results of more than 100 msec., with pairs in alphabetical order being faster than pairs in reverse order. No such order effects have been reported in simultaneous digit experiments for the task first used by Moyer and Landauer. Finally, the overall mean latencies and error rates in their study appear substantially higher than those found for comparable studies using digits. On the other hand, the relation of Lovelace and Snodgrass' study to Exp. III of the present article is not at all straightforward. In particular, no significant order effects were found in the present Exp. III. The relation of these two letter tasks needs to be explored.

The purpose of the present study was to investigate possible similarities between digit inequality judgment tasks and simple mental addition. If a comparison is made between the results of Exp. I and II of the present article and the results in Parkman and Groen (1971), a high degree of similarity between the two tasks is found. Substantial and comparable portions of mean reaction time variance in both tasks are described by a linear function of the Min. (The Min for simple addition problems corresponds to the minimum addend.) Also, order effects in the two tasks are either very small (12 msec. for the addition task) or nonexistant. However, the Min slope for addition was approximately 30 msec/unit on *S*s' first day as compared to the 10–20 msec/unit Min slopes found in the present study. This slope difference between the two tasks raises some very interesting possibilities for models of simple addition based upon the model proposed for the present digit inequality judgment task. For example, the counting model described by Parkman and Groen for simple addition assumes that somehow *S* finds the Max for the two addends and then computes the sum by setting a register equal to the Max and successively incrementing that register by an amount equal to the Min. It has been suggested (Parkman & Cavanagh, 1971) that *S* finds the Max in simple addition in the manner described by the model for the present Exp. I and that the later register incrementations are similar in temporal requirements to the counting behavior of the register in finding the Max. If this is true, then simple addition, which would involve two sets of register incrementations by the Min, should be described by a Min slope twice as large as that found for the present digit inequality tasks. Such a result is indicated by the findings of the present study in conjunction with those reported by Parkman and Groen, although further work along this line needs to be done.

REFERENCES

CLARK, H. H. Linguistic processes in deductive reasoning. *Psychological Review*, 1969, **76** 387–404.

DRAPER, N. R., & SMITH, H. *Applied regression analysis*. New York: Wiley, 1966.

FAIRBANK, B. A., JR. Experiments on the temporal aspects of number perception. *Dissertation Abstracts International*, 1969, **30** (1 B), 403.

GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.

LOVELACE, E. A., & SNODGRASS, R. D. Decision times for alphabetic order of letter pairs. *Journal of Experimental Psychology*, 1971, **88**, 258–264.

MOYER, R. S., & LANDAUER, T. K. Time required for judgments of numerical inequality. *Nature*, 1967, **215**, 1519–1520.

NUNNALLY, J. C. *Psychometric theory*. New York: McGraw-Hill, 1967.

PARKMAN, J. M., & CAVANAGH, J. P. A model for simple numerical processing. Paper presented at the Mathematical Psychology Meetings, Princeton, N.J., September 1971.

PARKMAN, J. M., & GROEN, G. J. Temporal aspects of simple addition and comparison. *Journal of Experimental Psychology*, 1971, **89**, 335–342.

RESTLE, F. Speed of adding and comparing numbers. *Journal of Experimental Psychology*, 1970, **83**, 274–278.

(Received April 24, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 206–211

COMPOUND AND SIMPLE RESPONSES IN PAIRED-ASSOCIATE LEARNING [1]

JOSEPH L. YOUNG [2] AND ROBERT L. SCHIFFER

State University of New York at Stony Brook

Isidore Newman School, New Orleans

Two groups of 30 undergraduates each were run in a paired-associate list-learning experiment. These groups differed only in the responses used; one had the simple responses 1, 2, 3, 4, and the other, the compound responses X1, X2, Y1, Y2. The simple response group learned more rapidly. The components of the compound response were not learned all or none in the standard sense, but the correctness of the components, both in the learning and in the precriterion phase, showed some (though not total) dependence on each other. Suggestions are given for a view of the learning processes involved that encompasses these results.

A difference commonly found between paired-associate learning experiments in the mathematical model tradition (cf. Atkinson, Bower, & Crothers, 1965; Atkinson & Crothers, 1964; Bower, 1967) and those coming from other traditions (cf. Underwood & Schultz, 1960) has been the choice of response terms. Those *E*s without an orientation toward mathematical models have typically used responses from extremely large sets (e.g., nonsense syllables or short words), while those with such an orientation have typically used simple responses from a small set identified beforehand to *S*s (e.g., single-digit numbers or letters of the alphabet).

Experiments using complex responses, such as nonsense syllables, require *S*s to learn which nonsense syllables are being used as responses as well as which stimulus each response "goes with." In such a situation, it is difficult for *E* to force *S*s to "guess," when they are unsure, and to work out the probability of their being correct when they do guess, requisites for analysis of the data with mathematical-model tools. The simple responses favored by mathematical modelers get rid of the problem of response learning and certainly make a guessing requirement for *S*s and à priori assessment of the guessing rate by *E* simple enough, but their use skirts completely the issues involved in response complexity. Compound responses (e.g., color *and* geometric form, letter *and* number) in which the number of response alternatives is small and fixed allow escape from the problems associated with use of extremely large response sets while permitting the retention of some complexity in the structure of the response.

Early work on compound responses was inconclusive. Bower extended his all-or-none conditioning model (Bower, 1961a) to compound responses (Bower, 1961b). The compound response model allows the learning of the two response components to proceed totally independently, to occur simultaneously, or to take on any of several kinds of partial dependence on each other, depending on the parameter values.

The only published experiment known to the authors that is concerned precisely with the study of compound responses in paired-associate learning was conducted by Crothers (1962).[3] His was a "miniature experiment," involving only two tests on each item, with responses being a colored

[1] This paper is based on an undergraduate research project performed by the second author under the direction of the first author. The project was aided by Grant MH 15577 from the National Institute of Mental Health.

[2] Requests for reprints should be sent to Joseph L. Young, Department of Psychology, State University of New York, Stony Brook, New York 11790.

[3] Another study, more similar to Group 1 of the present experiment than to Crothers', is referred to by Bower (1967), but with relatively little detail, so that the questions raised in the present paper are not answered. No data in the present study conflict with the analogous results reported by Bower, but much of the present analysis goes beyond that which he reported.

(red, green, or blue) geometrical form (circle or triangle). His conclusion was that the components were neither learned independently nor as a unit, but rather that getting one component correct increased the probability of getting the other correct. Subsequent work has made it clear that interpretation of Crothers' experiment is difficult for two reasons. First of all, the most stringent test of Bower's compound response model (Bower, 1961b) is its prediction of stationarity of response probability and independence of the correctness or incorrectness of responses on successive trials *for each component separately* prior to the last error *on that component* (Suppes & Ginsberg, 1963).[4] The necessary chi-square tests cannot be applied to Crothers' data, since these tests require a longer string of trials, such as occur in standard list-learning experiments. Second, it was not known then, as it is now (cf. Atkinson & Crothers, 1964; Bower, 1967), that Bower's (1961a) simple all-or-none model does not fit data with more than two response alternatives. Thus the Suppes and Ginsberg (1963) chi-square analysis, if it could be performed on data like Crothers', would presumably pick up violations of stationarity and independence prior to the last error, at least with the color component.

In addition to the question of how the response components are learned, a further question exists: How does the way compound responses are learned compare with the way simple responses are learned, when the same total number of response alternatives are employed?

The present experiment investigated these questions. One group of *S*s was given a list-learning task with two alternatives for each of two response components. A second group was given the same task, but with single-component responses involving four response alternatives.

[4] This follows from the fact that if a response is considered correct whenever a particular component is correct, Bower's model for compound responses (Bower, 1961b) reduces to his simple all-or-none model (Bower, 1961a), to which Suppes and Ginsberg's (1963) conclusions apply (cf. Bower, 1967).

METHOD

Subjects.—The *S*s were 60 students from an introductory psychology course at the State University of New York at Stony Brook who participated as part of a course requirement. They were assigned to two groups of 30, each group for one of the two experimental conditions. Two further *S*s were discarded because of failure to follow instructions.

Apparatus.—The *E* and *S* sat facing each other across a table, separated by a 30.48 × 91.44 cm. screen behind which *E* worked. Stimulus–response items were presented on 4 × 6 in. cards displayed above the screen. Each card was folded down the middle (into two 4 × 3 in. sections). The left-hand section (as seen by *S*) contained a three-letter consonant-vowel-consonant (CVC) nonsense syllable. The right-hand section section contained a letter and a number (Group 1) or just a number (Group 2).

Design.—Each *S* was required to learn 12 stimulus–response items. Twenty-four heterogeneous CVC syllables (none having the letters X or Y) were selected; these had meaningfulness values from 22% to 88% as tabled by Underwood and Schultz (1960). The 24 syllables were divided into two sets; each set was given to half of Group 1 and half of Group 2. For each group, the syllables were paired randomly with the responses, three stimuli to each of the four responses. Group 1's four responses were the compound responses, X1, X2, Y1, Y2, and Group 2's four responses were the simple responses 1, 2, 3, and 4.

Procedure.—Each *S*, on entering the room, was given standard list-learning-by-anticipation instructions. To familiarize *S* with procedures, the actual experiment was preceded by a practice task, involving syllables not used in the experiment proper as stimuli and the letters A through D and the digits 5–8 as responses. Both practice task and actual experiment were done by the standard anticipation method. The timing of presentations for the initial trial (i.e., cycle through the list of 12 items) was as follows: stimulus for 3 sec., stimulus and response for 3 sec., interpresentation interval of 1 sec., and intertrial interval of 30 sec. On subsequent trials, the timing was the same, except that the stimulus was presented until *S* responded. During the intertrial interval, *E* shuffled the stimulus–response cards in preparation for the next trial. Each *S* was run to a criterion of two consecutive completely correct trials.

RESULTS

The first item of interest in these data is the comparison of the overall speed of learning in the two groups. Beyond a shadow of a doubt, on any standard measure of speed of learning, Group 2, the group with the simple responses, learned faster than Group 1, the group with the compound responses. Two measures will

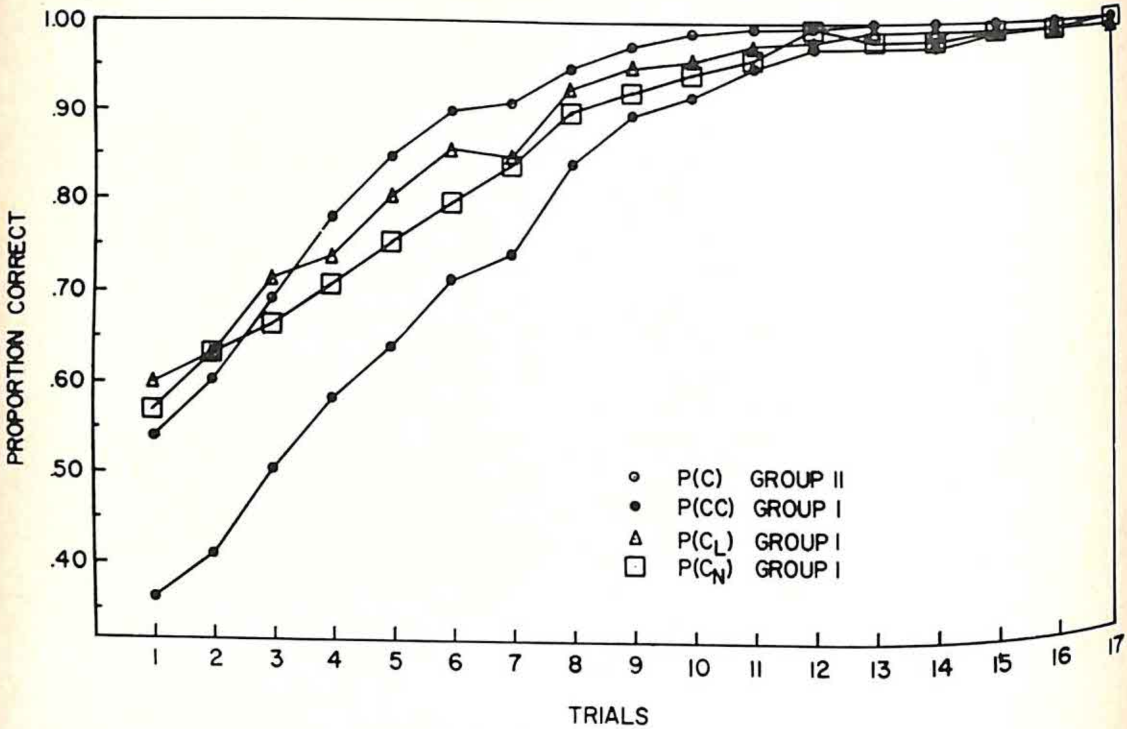

FIG. 1. Group learning curves. (P(C) refers to the proportion of correct responses in Group 2: $P(C_L)$, $P(C_N)$, and P(CC) refer, respectively, to the proportions of items in Group 1 with their letter components correct, their number components correct, and both components correct.)

suffice to prove the point: Group 1 took a mean of 12.57 trials to reach criterion, while Group 2 took a mean of only 8.33 trials, t (58) = 3.64, $p < .001$. Group 1 had a mean trial of last error (TLE), averaged across all S items, of 5.50, while Group 2's mean TLE was only 2.82, t (718) = 9.58, $p < .001$.

It might be argued that though the simple responses in Group 2 are learned more quickly than the total compound in Group 1, they might be learned more slowly, having four response alternatives, than the two-response-alternative number component (or letter component for that matter) of the compound. This, too, was not the case. The simple response of Group 2 (mean TLE = 2.82) was learned more quickly than the number component in Group 1 (mean TLE = 4.42, t (718) = 6.02, $p < .001$) and more quickly than the letter component in Group 1 (mean TLE = 3.61, t (718) = 2.79, $p < .01$).

Finally, it can be seen in Fig. 1, which gives the learning curves, that the overall proportion of correct responses in Group 2, P(C), was higher on all trials than the proportion of completely correct responses in Group 1, P(CC). On the early trials, Group 2's proportion of correct responses was below the proportion of correct responses on either component in Group 1 ($P(C_L)$ and $P(C_N)$), due in all likelihood to the differences in the number of response alternatives and hence in the guessing probability, but it climbed above the proportions for each component by the fourth trial and remained there.

The next issue to be considered is what, if anything, was learned in an all-or-none manner. As noted above, the fundamental properties of standard all-or-none learning are stationarity and independence prior to the last error.

The Suppes and Ginsberg (1963) chi-square tests allow the testing of the hypo-

TABLE 1

SUPPES AND GINSBERG CHI-SQUARE VALUES WITH ASSOCIATED DEGREES OF FREEDOM AND PROBABILITY VALUES

Measure	Group 1 Compounds			Group 1 Letters			Group 1 No.			Group 2		
	χ^2	*df*	*p*	χ^2	*df*	*p*	χ^2	*df*	*p*	χ^2	*df*	*p*
Independence[a]	11.96	1	<.001	14.87	1	.001	3.68	1	.055	8.27	1	.004
Forward stationarity	49.13	12	<.001	20.65	12	.06	12.55	12	.41	8.36	9	.49
Backward stationarity	16.88	10	.08	4.87	10	.90	11.98	10	.29	13.04	8	.11
Vincent quartile stationarity	—			6.39	3	.09	4.52	3	.21	15.91	3	.0012

Note.—Chi-square value for the Vincent quartile stationarity measure for Group 1 compounds was not computed, since significance of foward stationarity χ^2 made computation superfluous.
[a] Values for Group 1 computed on data prior to the last errors on both letters and numbers.

theses of stationarity and independence. The four tests, the results of which are given in Table 1, include a test for independence and three tests for stationarity, forward, backward, and Vincent quartile. These tests are all performed on data prior to the last error. The test for independence is a 2×2 contingency test to determine whether the correctness of the response on Trial n depends on the correctness of the response on Trial $n - 1$. The tests for stationarity are contingency tests to determine whether the proportion of correct responses differs in different parts of the data prior to the last error. The forward, backward, and Vincent quartile stationarity tests compare, respectively, the proportions of correct responses on successive trials indexed in the standard fashion, on successive trials indexed backwards starting with the trial just prior to the last error, and in the four Vincent quartiles of the data prior to the last error. A significant chi-square in the test for independence renders untenable the hypothesis of independence, while a significant chi-square in any of the three tests for stationarity renders untenable the hypothesis of stationarity.

The application of these tests, the results of which appear in Table 1, provides clear information. The responses in Group 2 and the total compound response in Group 1, as expected on the basis of their involving four response alternatives, were not learned in a standard all-or-none fashion; neither was the letter component in Group 1. The presence of at least one significant chi-square value in the corresponding columns of Table 1 indicates the failure of learning to be all or none. The hypothesis that the number component was learned in an all-or-none fashion is still tenable, but it is parsimonious to believe, on the basis of the chi-square for independence which fell just short of the level needed for significance at the .05 level, that in that case too, learning did not proceed in the standard all-or-none manner.

A third issue of concern is the dependence of the correctness of the components on each other in Group 1. First of all, on each trial, the proportion of correct responses on the letter component $P(C_L)$ was higher than the proportion of correct responses on the number component $P(C_N)$, which was in turn higher than the proportion of correct responses on the compound $P(CC)$ (see Fig. 1). Therefore, it is clear that there was not total and complete dependence of the correctness of the two components on each other; for if there were, $P(C_L)$, $P(C_N)$, and $P(CC)$ would be identical. Conversely though, since on each trial the product of $P(C_L)$ and $P(C_N)$ is less than $P(CC)$ and since their equality would be required for independence of the correctness of the two components, it is clear that some dependence existed.

Dependence existed even in data in which at least one error on each component was yet to come. Including only such data,

the proportion of correct letters was .68 when the number was also correct and only .53 when the number was incorrect. Similarly, the proportion of correct numbers was .63 when the letter was correct and .48 when it was not. These differences were highly significant, $\chi^2(1) = 14.22$, $p < .001$.

Even the speeds at which the two components of a response were learned were highly correlated. To avoid a possible influence of some sort of overall individual learning rate on such a correlation, for each *S*, a mean TLE for the compound response was computed. Then for each of the *S*'s items, the deviation of the TLE for the number and the TLE for the letter from *S*'s mean TLE were computed. These deviation scores were correlated for all the *S* items, yielding $r = .429$, equivalent to $t\ (358) = 8.99$, $p < .001$, a highly significant value; thus it is clear that TLEs of the components for a single item tend to occur together, be they early or late relative to *S*'s learning of the list.

Discussion

The clearest result in these data is that the learning rate for compound responses is lower than that for simple responses. From a simple information-processing point of view, this result seems surprising. In both cases, the responses were overlearned familiar symbols and carried 2 bits of information. In fact, the only real difference is that both bits were carried by the same symbol in the simple response case, while two symbols carried 1 bit each in the compound response case. It might seem according to this view that segmentation of the information into discrete bits might enhance rather than suppress learning in the compound response case, since *S* would have done for him the work of segmenting the information into bits, part of the supposedly efficient strategy of operating on 1 bit of information at a time. Clearly, humans do not operate on 1 bit of information at a time in this situation. In fact, the existence of a second symbol, the letter, along with the 1-bit number in the compound response, seems to have retarded processing of the number below the level experienced by the 2-bit simple response number. The data suggest, then, that each symbol is processed singly, regardless of its information content, and that the existence of a second symbol retards processing of the first. Incidentally, the latter interpretation is bolstered by unpublished experiments done by W. K. Estes and his students some years ago at Indiana University. (W. K. Estes, personal communication, July 1970.) They compared paired-associate learning of a standard sort with a situation in which there were two correct responses, either of which would suffice on any test, and found that *S*s with two correct responses did far worse than those with the standard task; the existence of a second symbol retarded processing of the first, as in the present experiment. The precise reason for this retardation is not clear, though it may relate to channel overload or like factors.

The results of the internal analysis of the compound response group are certainly consonant with Crothers' (1962) conclusion of a partial dependence of the correctness of the two components on each other. However, the fact that the individual components seem not to have been learned in the standard all-or-none fashion is troublesome. What seems to be indicated is a view in accord with recent mathematical model formulations involving short-term memory states. What may well have happened is that learning of each component was all or none, but not in the standard fashion (Bower, 1961a, 1961b). First of all, the dependence of a correct response on one component on a correct response on the other component prior to the last errors on both indicates that short-term memory, at least, seems to have been for the compound and not for the individual components. This indication arrives from the property, common to models involving short-term memory states, that correct responses prior to the last error are either guesses or consequences of short-term memory. Second, the correlation of TLE for the two components seems to indicate that though long-term learning may be all or none, its probability of occurrence may be different on one component if the other component is or is not learned, as originally postulated by Bower (1961b). Clearly, the details of such a model, adding a short-term memory state for the compound to Bower's notions, would have to be worked out in more specificity.

References

Atkinson, R. C., Bower, G. H., & Crothers, E. J. *An introduction to mathematical learning theory.* New York: Wiley, 1965.

Atkinson, R. C., & Crothers, E. J. A comparison of paired-associate learning models having different acquisition and retention axioms. *Journal of Mathematical Psychology*, 1964, 1, 285–315.

Bower, G. H. Application of a model to paired-associate learning. *Psychometrika*, 1961, 26, 255–280. (a)

Bower, G. H. Application of the all-or-none conditioning model to the learning of compound responses. (Tech. Rep. No. 37) Palo Alto, Calif: Stanford University, Institute for Mathematical Studies in the Social Sciences, 1961. (b)

Bower, G. H. Notes on a descriptive theory of memory. In D. P. Kimble (Ed.), *The organization of recall: Proceedings of the second conference on learning, remembering, and forgetting*. Vol. 2. New York: New York Academy of Sciences, 1967.

Crothers, E. J. Paired-associate learning with compound responses. *Journal of Verbal Learning and Verbal Behavior*, 1962, 1, 66–70.

Suppes, P., & Ginsberg, R. A fundamental property of all-or-none models: Binomial distribution of responses prior to conditioning, with application to concept formation in children. *Psychological Review*, 1963, **70**, 139–161.

Underwood, B. J., & Schultz, R. W. *Meaningfulness and verbal learning*. Philadelphia: Lippincott, 1960.

(Received February 4, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 212–214

RETROACTIVE INHIBITION IN THE A-B, A-D PARADIGM AS MEASURED BY A MULTIPLE-CHOICE TEST [1]

COLEMAN T. MERRYMAN [2]

University of Texas at Austin

In 1969, L. Postman and K. Stark failed to find retroactive inhibition (RI) in the A-B, A-D paradigm when performance was measured by a multiple-choice recognition test. They concluded that specific A-B associations are highly resistant to unlearning. An alternative explanation, tested in the present study, is that A-B associations do undergo RI, but performance on the multiple-choice test is mediated by backward associations. The pairing of B terms as well as A terms with new items during interpolated learning (a procedure producing RI of both backward and forward associations when measured by a recall test) also produced RI of A-B when measured by a multiple-choice test. The results were discussed in terms of their relation to the question of differences between recognition and recall.

One of the best established phenomena in the area of verbal learning is the decrement in recall of a response to a given stimulus following the pairing of the same stimulus with an unrelated response. That is, paired-associate (PA) training on A-D following training on A-B produces retroactive inhibition (RI) of A-B, the inhibition being measured relative to an A-B, C-D control. Until recently, this RI had been theoretically interpreted as arising from two sources, competition between the B and D responses and unlearning of specific A-B associations during A-D training (Melton & Irwin, 1940).

The unlearning mechanism of RI has been questioned by Postman, Stark, and Fraser (1968) and Postman and Stark (1969). These investigators presented data which they interpreted as indicating that specific A-B associations are not unlearned during A-D training. Rather, their data indicated that the whole set of B response terms becomes suppressed during A-D learning. Consequently, on the test for A-B retention, the B items are not available, and the A-B associations (although still quite intact) are not manifested in data. In other words, the associate of Stimulus A is a B item which is not available for recall.

A simple experiment provided considerable support for this suppression hypothesis. If, in the test to determine the strength of A-B following interpolated learning, the B terms are made available by means of a multiple-choice test, then no RI should be evidenced. Postman and Stark (1969) reported just such a finding. Although a typical A-B recall test revealed considerable RI in the A-B, A-D paradigm relative to A-B, C-D, performance on a multiple-choice test was essentially perfect in both paradigms.

An alternative explanation of the Postman-Stark data is presented in the present paper, an explanation based on *S*'s hypothesized use of backward associations to mediate performance on the multiple-choice test. It is well established that *S*s concomitantly learn B-A as they learn an A-B paired-associate list. Furthermore, there is ample evidence from Johnston (1967), Merryman (1969), Petrich (1970), and Wolford (1969) that forward and backward associations can undergo RI (as measured by the conventional recall method) independently of each other. If A-D train-

[1] The author gratefully acknowledges the contributions of Arthur Kauffman, who prepared the stimulus materials; of Arthur Kauffman, Gordon McKinley, and Chris Myers who ran *S*s; of the above and Marieke Wempe who assisted in data analysis; and of an anonymous reviewer who suggested several improvements in the manuscript.

A paper based on Exp. I was presented at the meetings of the Midwestern Psychological Association, Cincinnati, May 1970.

[2] Requests for reprints should be sent to Coleman T. Merryman, 211 Mezes Hall, Department of Psychology, University of Texas, Austin, Texas 78712.

ing produces unlearning of specific A-B but not of B-A associations, then correct responding on a multiple-choice test of A-B retention can be mediated by the still intact backward associations.

If the explanation in terms of backward associations is correct, then the pairing of first-list response terms with new items (B-E) during A-D training should produce unlearning of both forward and backward associations; therefore, a multiple-choice test of A-B retention should reveal considerable RI relative to both the A-B, A-D and the A-B, C-D paradigms, which should not differ from each other. These predictions were tested in two experiments.

Method

Design and procedure.—Each *S*, in a within-*S*s design, learned a PA list (A-B) by the study-recall method to a criterion of one perfect recitation or 15 trials, whichever came first. Materials were presented via a Lafayette memory drum at a 2-sec. rate during study and recall cycles. Study cycle-recall cycle and recall cycle-study cycle intervals were 2 and 6 sec., respectively.

Upon reaching criterion, *S*s received brief instructions concerning List 2, which was then presented in the same manner as List 1. Training on List 2 continued to the same criterion as in List 1 in Exp. I, and for 15 trials for all *S*s in Exp. II.

For one-third of the List 1 pairs, the stimuli were paired with new responses in List 2; those pairs correspond to the A-B, A-D paradigm. For another third of the List 1 pairs, both the stimuli and responses were paired with new responses in List 2; they can be referred to as the A-B, A-D & B-E paradigm. (That is, in the A-D & B-E paradigm in List 2, both the List 1 stimuli and the List 1 responses were used as stimuli and were paired with new response terms.) For the remaining third of the List 1 pairs, neither the stimuli nor responses appeared in List 2, so those pairs served as controls. The set of List 1 pairs that corresponded to each paradigm was counterbalanced.

Following criterion on List 2, a one-trial unpaced multiple-choice test of retention of the List 1 pairs was administered. Index cards were presented one at a time in random order. On the left of each card was one A term and on the right were four B terms, the B terms all being the response members of pairs from the same paradigm as the presented A term. The *S* was asked to choose the correct B term from among the four presented.

Materials.—Each list consisted of 12 pairs. The A and E terms were single letters, and the B and D terms were four-letter monosyllabic nouns with Thorndike-Lorge ratings of A and AA. Within a list, no two nouns began with the same letter and no word began with any of the single letters used.

Subjects.—For Exp. I and II, respectively, 45 and 48 female undergraduates at the University of Texas at Austin served as *S*s.

TABLE 1

PROPORTION OF CORRECT RESPONSES (CORRECTED FOR GUESSING) ON THE MULTIPLE-CHOICE TEST

Exp.	Paradigm		
	A-B, A-D & B-E	A-B, A-D	Control
I	.76	.94	.97
II	.80	.96	.97

Results

First-list learning—Pairs subsequently designated to the three paradigms did not differ significantly ($F\text{s} < 1$) in first-list learning in either experiment. Mean errors to criterion were 16.2, 15.4, and 16.0 in Exp. I and 18.3, 18.0, and 18.7 in Exp. II for the A-B, A-D, the A-B, A-D & B-E, and the control paradigms, respectively.

Second-list learning.—The three kinds of pairs in the second list were (*a*) B-E, (*b*) A-D corresponding to those pairs which also received B-E (denoted $A\text{-}D_{B\text{-}E}$), and (*c*) A-D corresponding to those pairs which did not receive B-E (denoted A-D). Mean errors were 11.1, 14.4, and 14.6 in Exp. I and 10.8, 16.2, and 15.5 in Exp. II for the B-E, $A\text{-}D_{B\text{-}E}$ and A-D pairs, respectively. These differences are significant in both experiments, $F\ (2, 88) = 10.3$ and $F\ (2, 94) = 16.5$ for Exp. I and II, respectively, ($p\text{s} < .01$). Neuman-Keuls tests revealed that both the $A\text{-}D_{B\text{-}E}$ and A-D pairs differed from the B-E pairs in both experiments ($p\text{s} < .01$) but not from each other ($p\text{s} > .05$).

Multiple-choice test.—Of primary importance are the results of the multiple-choice test, and these results are very clear. The proportion of correct responses (corrected for guessing) for the three paradigms are presented in Table 1 for both experiments. Performance in the A-B, A-D and the control paradigms was virtually perfect, a

result compatible with that of Postman and Stark (1969); but performance following A-D & B-E List 2 training was greatly reduced. Analyses of variance confirm the obvious, F (2, 88) = 17.4 and F (2, 94) = 17.7 for Exp. I and II, respectively (ps < .01). In both experiments, Neuman-Keuls pair-wise tests show performance in the A-B, A-D & B-E paradigm to be inferior to each of the other paradigms (ps < .01). The difference between the latter two paradigms did not even approach significance.

Discussion

The results support the hypothesis that *S*s make use of backward associations formed during first-list learning to mediate performance on a multiple-choice test and that specific A-B associations are unlearned during A-D training. The results are not compatible with a suppression hypothesis, since it is difficult to imagine how the selector mechanism would suppress responses in the A-B, A-D & B-E paradigm, but not in the A-B, A-D paradigm. Additional support for the unlearning of specific A-B associations has been provided by Weaver and McCann (1970) employing different experimental procedures.

The results of the present experiments call into question one source of evidence used to support the hypothesis of separate mechanisms for recall and recognition. That evidence is that RI occurs in recall but not in recognition tasks (see, e.g., Kintsch, 1970). The present study demonstrates that RI is obtained on a multiple-choice recognition test when the strength of the mediating backward associations is reduced.

References

Johnston, W. A. S-R, R-S independence and the interference potency of latent R-S associations. *Journal of Experimental Psychology*, 1967, **74**, 511–516.

Kintsch, W. Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

Melton, A. W., & Irwin, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *American Journal of Psychology*, 1940, **53**, 173–203.

Merryman, C. T. Effects of strategies on associative symmetry. (Indiana Mathematical Psychology Program Rep. No. 69-6) Bloomington, Ind.: University of Indiana, 1969.

Petrich, J. A. S-R and R-S unlearning as a function of transfer paradigms. *Journal of Experimental Psychology*, 1970, **83**, 19–24.

Postman, L., & Stark, K. Role of response availability in transfer and interference. *Journal of Experimental Psychology*, 1969, **79**, 168–177.

Postman, L., Stark, K., & Fraser, J. Temporal changes in interference. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 672–694.

Weaver, G. E., & McCann, R. L., Jr. A mixed-list comparison of retroactive inhibition in the A-B, A-C and A-B, D-C paradigms. Paper presented at the Midwestern Psychological Association, Cincinnati, May 1970.

Wolford, G. L., II. Functions of multiple directional associations in verbal learning. Unpublished doctoral dissertation, Stanford University 1969.

(Received February 18, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 215–219

MEMORY SEARCH THROUGH CATEGORIES OF VARYING SIZE[1]

WILLIAM METLAY,[2] ALBERT HANDLEY

Hofstra University

AND IRA T. KAPLAN

New York University Medical Center

Each of three groups of Ss memorized a different list of names: signs of the zodiac, United States Presidents, and states of the United States of America. After memorizing the list, S was required to write all the names that contained a specified letter. Three letters were specified for each category, and each letter designated a different number of target names within the category. In every case, the cumulative distribution of response times was described by an exponential curve. The parameters of the fitted curve estimated the number of targets that S could recall, and the proportional rate at which this limit was approached. The proportional response rate was found to be independent of the number of targets, but varied inversely with the total number of names in the category. These results were expalined in terms of a random search model of verbal recall.

When S lists all the words he can recall from a specified category of information stored in his memory, his response rate decreases as he exhausts his supply of recallable words. Bousfield and Sedgewick (1944) showed that the cumulative distribution of recall times can be described by the negatively accelerated exponential curve

$$F = T\,(1 - e^{-pt}), \qquad [1]$$

where F is the number of responses given by time t, T is the total number of words recallable, and p is an index of the relative response rate, i.e., the larger the value of p the more rapidly the curve approaches its upper limit, T.

A priori, there is no mathematical reason why p and T should be related, but experimental studies of verbal recall have shown an inverse relation between them (Johnson, Johnson, & Mark 1951; Rogers, 1956). Subsequent studies (Indow & Togano, 1970; Kaplan, Carvellas, & Metlay, 1969, 1971) confirmed this inverse relation and explained it in terms of a search model in which S scans a memory display for target words. The targets are selected randomly from a larger number of items on the display, and the relative response rate varies directly with S's scanning rate, k, and inversely with the total number of items, N, i.e.,

$$p = \frac{k}{N}. \qquad [2]$$

According to this explanation, the inverse relation between p and T observed in the above experiments occurred because the manipulations that increased the number of target words also increased the number of items in the search set.

The present experiment was designed to test the hypothesis that p depends directly on N rather than on T by manipulating the number of target words and the total number of items in the search set independently. The procedure was similar to one described by Anderson (1969). Specifically, Ss memorized a list of items, which they then searched for target words that contained a specified letter. The number of items, N, and the number of targets, T, in the list were varied in order to observe the effects of these variables on the proportional rate, p, at which the targets were found.

METHOD

Subjects.—The Ss were 34 students in three experimental psychology classes of 12, 12, and 10

[1] This investigation was supported by United States Public Health Service Research Grant EY–00384 from the National Eye Institute, and by Hofstra University Grant 0580.

[2] Requests for reprints should be sent to William Metlay, Department of Psychology, Hofstra University, Hempstead, New York 11550.

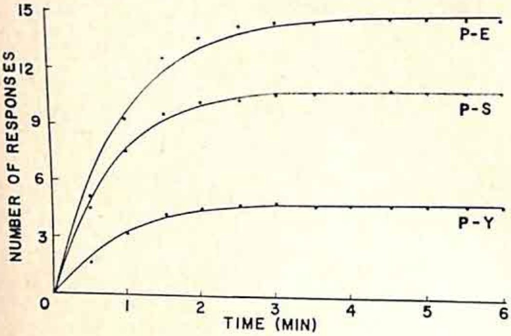

FIG. 1. Cumulative number of target names found as a function of time for the three Presidents conditions. (For the upper, middle, and lower curves, the targets were names containing the letters *e*, *s*, and *y*, respectively.)

students at Hofstra University. Each class was tested as a group with a different category of names.

Procedure.—In the first part of the experiment, *S*s studied the names in one of the following three categories: the 12 signs of the zodiac, the 32 different surnames of United States Presidents, or the 50 states of the United States. When *S*s entered the classroom, the names in each category were already printed in random order and in random locations, i.e., not in rows or columns, on the blackboard in front of the room. Then *E* read all the names in the category aloud in random order. As each name was read, *S*s found it on the board, read it aloud, and spelled it in unison. After this oral rehearsal was completed, *S*s were instructed to write on sheets of paper all the names they could recall in any order, without looking at the blackboard. The procedure of oral rehearsal followed by written recall was repeated until every *S* learned most or all of the names of his category. For the zodiac, Presidents, and states categories this procedure was repeated three, five, and three times, respectively. The importance of correct spelling was emphasized throughout.

In the second part of the experiment, *S* was required to recall certain names from the category he had just learned, after the names were erased from the blackboard. The target names were designated by specifying one letter of the alphabet that was common to all of them, e.g., in the zodiac conditions the letter *n* designated Gemini, Cancer, and Capricorn. At the beginning of this part of the experiment each *S* received three sheets of paper face down. On instruction, *S* turned the first sheet over to see the letter that indicated the names he was to write on that page. The names were written in a column one beneath the other. To mark time on his answer sheet, *S* drew a line under his last response at a signal that *E* gave every 30 sec. The *S*s responded for 5, 6, and 8 min. in the zodiac, Presidents, and states conditions, respectively. After the first sheet was collected, this procedure was repeated with different letters for the second and third sheets. The number of names designated by a given letter varied: In the zodiac conditions, the letter *n* indicated 3 names, *s*, 6 names, and *i*, 9 names; for Presidents, *y* indicated 5 names, *s*, 11 names, and *e*, 15 names; for states, *k* indicated 9 names, *l*, 15 names, and *a*, 36 names. The three letters that each *S* was tested with were presented to different *S*s in counterbalanced order.

RESULTS

After the last rehearsal of the names in a category, *S*s recalled all 12 signs of the zodiac in 3 min., all 32 Presidents' surnames in 4 min., and an average of 48 states in 11 min.

When *S* was required to list the names that contained a specified letter, the number of responses was limited by the number of target names he had learned. Figure 1 shows how the cumulative response frequency approached the number of targets in the three Presidents conditions. Each point represents the average number of responses for 12 *S*s. The curves are graphs of Equation 1 which were fitted by an optimization procedure that minimized the sum of deviations squared.[3] There were 5, 11, and 15 targets in these three conditions and the response frequencies nearly reached these levels within 6 min. The estimated limits, *T*, of the fitted curves were approximately equal to the actual number of designated names in each condition, i.e., 4.97, 10.82, and 14.74. Similar results were obtained in the zodiac conditions, where the actual numbers of designated names were 3, 6, and 9, and the corresponding *T* values were 2.90, 5.80, and 8.44. In the states conditions, the estimated *T* values of 7.66, 13.66, and 32.85, were somewhat less than the 9, 15, and 36 actual target names, which is to be expected, since *S*s were unable to list all the states on the last recall trial.

It is difficult to appreciate differences in relative rate, *p*, by visual inspection of the cumulative frequency distributions when these distributions rise to different limits. Differences in relative rate can be made

[3] The authors are grateful to M. Goldberg, W. Kern, and J. Neiditch for writing the computer program that calculated the fitted curves.

more apparent, however, by plotting cumulative frequencies as percentages of their estimated T values. This has been done in Fig. 2, where the cumulative percent of responses is plotted as a function of time for curves with three different p values. Estimates of p were obtained for all conditions of the experiment by fitting Equation 1 to the data. To show the separation between the curves more clearly, only the first 5 min. of responding have been plotted, since after 5 min. the three curves nearly coincide. The zodiac curve, which has the relatively high p value of 2.20/min rises more steeply and reaches 80% of its limiting value in about .75 min. The Presidents curve, with p equal to 1.05/min, takes 1.5 min. to reach the 80% level, whereas the states distribution, with p equal to .61/min, has the shallowest slope and requires 3 min. to reach the same level.

The relative rate at which each cumulative distribution approached its limit was a function of the total number of names in the category. In the zodiac condition, where there were 12 names in the category, the p values were 3.83, 1.99, and 2.20 per minute for 3, 6, and 9 targets. The p values were lower for the Presidents category, which had 32 names: for 5, 11, and 15 targets, p was 1.10, 1.23, and 1.05 per minute. In the states category, which had 50 names, p was .77, .61, and .42 per minute in the 9-, 15-, and 36-target conditions, respectively. Thus, there was no systematic variation in p with the number of targets, but p did decrease systematically as the number of names in the category increased.

The hypothesis that p varies inversely with N can be tested by plotting $\log p$ against $\log N$. Taking the logarithm of Equation 2 yields

$$\log p = \log k - \log N. \qquad [3]$$

Assuming that the scanning rate, k, was constant across the conditions of this experiment, a graph of $\log p$ versus $\log N$ should be a straight line with intercept equal to $\log k$ and a slope of -1. The magnitude of N can be assumed to equal the number of items that S was able to

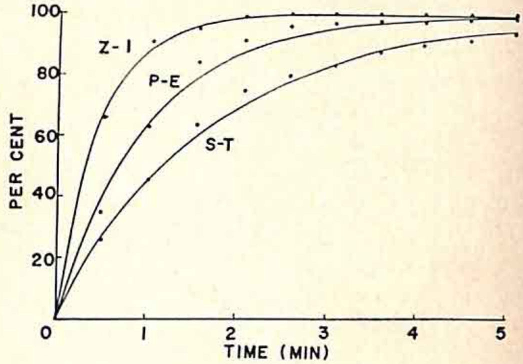

FIG. 2. Cumulative response frequency as a percentage of the limiting number of targets recallable. (The three curves represent the results obtained searching for zodiac names with the letter t, Presidents' names with the letter e, and states with the letter t.)

list on the last recall test before he searched for target words: for the zodiac category, $N = 12$; for Presidents, $N = 32$; and for states $N = 48$. In Fig. 3, the p values obtained for all nine conditions are plotted against their corresponding N values on log-log coordinates. The straight line, fitted by the method of least squares, has the equation $\log p = 1.53 - 1.02 \log N$. The slope of -1.02 is in excellent agreement with the expected theoretical value, and the antilog of the intercept indicates that S searched his memory at the rate of 34 items/min.

Discussion

The rate at which S retrieved target words from a category stored in his memory depended upon the number of items in the category: the larger the number of items, the slower the recall. This result is consistent with Shiffrin and Atkinson's (1969) description of recall as a search process in which a succession of memory codes is examined one after another and the information recovered is either rejected or accepted. Response latency is an increasing function of the number of images examined before the desired information is found. It follows that when S has to search through a larger set of stored codes, it takes him longer to find the desired item.

The random search model is a simplified description of verbal recall. It assumes that every item in the search set has an equal chance

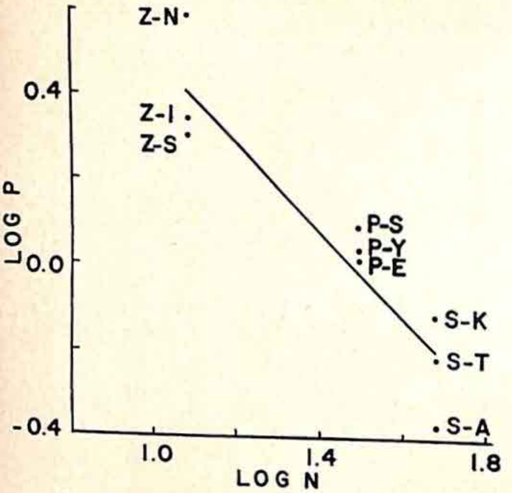

FIG. 3. Log proportional response rate, p, per minute versus log total number of names recallable in a category, N. (The nine data points were obtained searching for zodiac names with the letter n, i, or s; Presidents with s, y, or e; and states with k, t, or a.)

of being selected for examination at any moment, and that the time spent making a response, i.e., motor time, is negligible. Neither of these assumptions is entirely true. It has been shown that the response probabilities of the names in a category are unequal (Kaplan & Carvellas, 1969) and that the probabilities of successive responses in free recall of a memorized set of words are not independent (Tulving, 1962b). There is evidence for both nonrandomness and motor time in the present experiment, although these factors had little effect on the exponential distribution of recall times and on the dependence of relative response rate upon category size.

If recall were completely random, there would be no consistency in the order in which the target words were produced. But analysis of the data indicated a tendency for different *S*s to recall words in the same order. This tendency was shown by calculating, across *S*s, the average rank order of emission for the responses in a category. In the six-target zodiac condition, for example, the average order of emission was Sagittarius, Aquarius, Scorpio, Aries, Taurus, and Pisces, with mean ranks of 1.9, 3.4, 3.5, 3.6, 4.0, and 4.1. Two effects are apparent in this example, which were typical of the other conditions as well: First, the words that started with the target letter, *s*, were produced early in the series. This result is consistent with the observation that providing the initial letters of words aids recall (Earhard, 1967). Second, many words had nearly tied ranks, which indicates that there was no consistent sequence in their recall order.

It might be imagined that *S*s tried to recall the words in alphabetic order, since Tulving (1962a) showed that organizing words in alphabetic order facilitates retrieval. In fact, there was a correlation between the average order of emission and alphabetic order. This correlation was only $r = .03$ in the above example, but went as high as $r = .66$, with an average of $r = .30$ for all the conditions of the experiment. Indow and Togano (1970) demonstrated that systematic memory search produces a linear rather than an exponential distribution of recall times. The degree of organization in the present study was sufficiently small, however, in that the cumulative distributions were nearly exponential in form.

The effect of motor time would be to slow down the rate at which *S* finds targets, thereby decreasing p. The more targets in a condition, the larger the value of p. In nearly all the experimental conditions, *S* wrote only a few words per minute after the first minute of responding, so that most of his time was spent in memory search, selecting and testing items, rather than in writing targets. Only in the states condition with 36 targets were more than 4 targets/min produced after the first minute. In this condition, the relative rate was indeed the lowest in the states category. Among the nine conditions, there were two cases in which the largest number of targets was associated with the lowest p value (P-E and S-A in Fig. 3) and two cases in which the fewest targets were associated with the highest p value in a category (Z-N and S-K in Fig. 3). The influence of motor time is clearly visible in the data, but it is much smaller than the effect of category size. Thus, in spite of its simplifying assumptions the random search model described the cumulative distribution of recall times and, more significantly, predicted how the relative response rate varied with the number of items in a category.

REFERENCES

ANDERSON, N. H. A search task. In J. F. Voss (Ed.), *Approaches to thought*. Columbus, O.: Charles E. Merrill, 1969.

BOUSFIELD, W. A., & SEDGEWICK, C. H. W. An analysis of restricted associative responses. *Journal of General Psychology*, 1944, **30**, 149–165.

EARHARD, M. The facilitation of memorization by alphabetic instructions. *Canadian Journal of Psychology*, 1967, **21**, 15–24.

INDOW, T., & TOGANO, K. On retrieving sequence from long-term memory. *Psychological Review*, 1970, **77**, 317–331.

JOHNSON, D. M., JOHNSON, R. C., & MARK, A. L. A mathematical analysis of verbal fluency. *Journal of General Psychology*, 1951, **44**, 121–128.

KAPLAN, I. T., & CARVELLAS, T. Response probability in verbal recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 344–349.

KAPLAN, I. T., CARVELLAS, T., & METLAY, W. Searching for words in letter sets of varying size. *Journal of Experimental Psychology*, 1969, **82**, 377–380.

KAPLAN, I. T., CARVELLAS, T., & METLAY, W. Effect of context on verbal recall. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 207–212.

ROGERS, C. A. The orectic relations of mathematically derived fluency scores. *Journal of General Psychology*, 1956, **55**, 85–102.

SHIFFRIN, R. M., & ATKINSON, R. C. Storage and retrieval processes in long-term memory. *Psychological Review*, 1969, **76**, 179–193.

TULVING, E. The effect of alphabetical subjective organization on memorizing unrelated words. *Canadian Journal of Psychology*, 1962, **16**, 185–191. (a).

TULVING, E. Subjective organization in free recall of "unrelated" words. *Psychological Review*, 1962, **69**, 344–354. (b)

(Received March 15, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 220–226

COMPARISON OF RECOGNITION AND RECALL IN A CONTINUOUS MEMORY TASK

GEOFFREY R. LOFTUS[1]

Stanford University

Differences between recall and recognition performance may be due at least in part to differences in the way information is stored. This possibility was explored in a paired-associate learning task by varying *S*'s knowledge at the time of study of how he would be tested on a particular stimulus–response pair. It was found that when *S* knew how he would be tested, his performance was better on recall but worse on recognition than when he did not know how he would be tested. These results were interpreted as support for the assertion that differences in storage processes partially account for recall-recognition performance differences. A model which postulates a distinction between short-term and long-term memory provided an excellent fit to the data and suggested possible storage strategies for recall and recognition.

Studies comparing recognition and recall have found performance differences between the two. When raw scores are considered, recognition is generally superior to recall (Freund, Brelsford, & Atkinson, 1969; MacDougall, 1904; Postman, 1950; Postman, Jenkins, & Postman, 1948. It can be assumed that this superiority is due, at least in part, to differences in processes taking place at the time of *retrieval*; in particular, the probability of guessing the correct response is higher in recognition than in recall. The major purpose of the present experiment is to investigate the extent to which recall-recognition differences in a paired-associate task may be attributed to differences in *S*'s method of *storing* information.

A convenient way to carry out such an investigation is to vary *S*'s knowledge at the time of study of how he will subsequently be tested on some item. If he lacks this knowledge, *S* will be forced to study all items in the same way, and recall-recognition differences on such items must be attributable to retrieval processes alone. Having the knowledge, on the other hand, enables *S* to store information differentially for recall and for recognition if he so desires. In the present experiment there are three conditions under which *S* may study an item: (*a*) He may study with the knowledge that he will be tested by recall; items in this condition are *recall* (Re) items. (*b*) He may study knowing that he will be tested by recognition; items in this condition are *recognition* (Ro) items. (*c*) Finally, *S* may study knowing that he will be tested either by recognition or recall but not knowing which; items in this condition tested by recognition are *recognition mixed* (RoM) items, and those tested by recall are *recall mixed* (ReM) items. Consider *S*'s performance when tested on the various types of items: If Ro does not differ from RoM and Re does not differ from ReM, the implication is that storage differences are not necessary to account for recall-recognition differences. To illustrate why this is so, consider a storage operator, $\mathbf{S}$, which we use to represent various rehearsal schemes, mnemonic coding and, in general, any device used by *S* to store information for subsequent retrieval. The application of $\mathbf{S}$ to an input $\mathbf{I}$ represents information stored in memory, $\mathbf{S}(\mathbf{I})$. Further, a retrieval operator, $\mathbf{R}$, represents memory search, processing of stored information, and output of response at the time of test. Application of $\mathbf{R}$ to stored information produces output p, which is the probability of a correct response; i.e., $p = \mathbf{R}(\mathbf{S}(\mathbf{I}))$. For each of the three study conditions, there is a storage operator; these are designated $\mathbf{S}_{re}$, $\mathbf{S}_{ro}$, and $\mathbf{S}_{m}$ for recall, recognition, and mixed. Similarly, for the

[1] Requests for reprints should be sent to Geoffrey R. Loftus, now at New York University, Department of Psychology, 707 Broadway, New York, New York 10003.

two test conditions there are retrieval operators, $\mathbf{R}_{re}$ and $\mathbf{R}_{ro}$ for recall and recognition. Thus performance of the four types of output may be represented by:

$$\begin{aligned} \text{Ro—}p_{\text{Ro}} &= \mathbf{R}_{ro}(\mathbf{S}_{ro}(\mathbf{I})) \\ \text{RoM—}p_{\text{RoM}} &= \mathbf{R}_{ro}(\mathbf{S}_{m}(\mathbf{I})) \\ \text{Re}M\text{—}p_{\text{ReM}} &= \mathbf{R}_{re}(\mathbf{S}_{m}(\mathbf{I})) \\ \text{Re—}p_{\text{Re}} &= \mathbf{R}_{re}(\mathbf{S}_{re}(\mathbf{I})) \end{aligned}$$

If $p_{\text{Ro}} = p_{\text{RoM}}$ and if $p_{\text{Re}} = p_{\text{ReM}}$, then we may conclude that $\mathbf{S}_{ro}$ is the same as $\mathbf{S}_m$ and that $\mathbf{S}_{re}$ is the same as $\mathbf{S}_m$. Thus $\mathbf{S}_{re}$ and $\mathbf{S}_{ro}$ are identical, i.e., storage is the same for recall and recognition.

This paradigm was used by Freund et al. (1969) using paired-associate lists learned by the anticipation method. Their findings suggest that recall-recognition differences are due to retrieval alone; for both recall and recognition, varying *S*'s opportunity to store differentially did not lead to performance differences. Their design, however, may have been biased against finding storage differences for several reasons. First, their two-alternative forced-choice "recognition" test did not differ greatly from their nine-alternative forced-choice "recall" test. Second, and more important, *S*s were given all conditions within a single experimental session. The *S* would first study an item with the information that it would be tested by recall. This was followed by the study of a "recognition" item and then by a "test unknown" item. The cycle would then begin again with another "recall" item. To store differentially, *S* would have to change his method of study on a moment-to-moment basis. It is quite possible that this strategy of constant switching is too difficult to use.

The present study was similar in design to the Freund et al. (1969) experiment but it had two important changes. (*a*) A yes–no test was used for recognition. The response set consisted of the letters of the alphabet; thus "recall" was essentially from 26 alternatives. It was expected that these two tests would be more effective in demonstrating recall versus recognition differences than the 2- and 9-alternative forced-choice tests used by Freund et al. (1969). (*b*) Within an experimental session, *S* was in only one study condition: recall, recognition, or mixed. A single storage strategy could therefore be used throughout a session without eliminating the possibility of storing differentially as a function of condition.

Method

Subjects.—The *S*s were eight female graduate students who received \$2.00 for each session.

Apparatus.—The control functions of the experiment were performed by a computer program running on an on-line, modified, PDP-1d computer manufactured by Digital Equipment Corporation. The program operated on a time-sharing basis to drive eight KSR-33 teletypes which were situated in a single, windowless, soundproofed room in another building. Each *S* sat at a teletype equipped with a standard keyboard and a continuous roll of paper, masked in such a way that a horizontal strip about ½ in. wide was all that would be seen at a given time.

Procedure.—The stimuli were the digits 1–9, and the responses were the 26 letters of the alphabet. The *S*s served in one experimental session per day which took approximately 60 min. A within-*S*s design was used; each *S* was in each of the three experimental conditions five times and was in only one condition per session. The experimental conditions for each *S* were randomized over sessions, with the restrictions that she would be in all three conditions over a three-session block and would never be in the same condition for two sessions in a row. Each *S* had one initial practice session in the mixed condition.

A session started when *S* hit a code key on the teletype. The teletype would then print out what type of condition *S* was in for that day: recall, recognition, or mixed. A continuous task was employed: the nine stimuli were initially paired with randomly selected responses, and then the test-study trials which constituted the bulk of the session began. A stimulus to be tested was chosen randomly; in the mixed condition the type of test (recall or recognition) was determined randomly. The *S* had 30 sec. to recall the correct response and type in the appropriate letter in a recall test or, in a recognition test, had 30 sec. to type one of two special keys marked "yes" or "no." In recognition tests, the program randomly decided whether to present the correct response or a foil with the stimulus; if a foil was chosen, it was picked randomly from the 25 incorrect letters. If *S* was correct, feedback in the form of "+ + +" was printed out; if she was incorrect, "— — —" was printed out. The stimulus would then be re-paired with another response chosen randomly with the restriction that it could not be the one which had just been the correct response. The *S* was given 2 sec. to study the new pairing after which two carriage returns caused all typed material to disappear behind a mask, and *S*

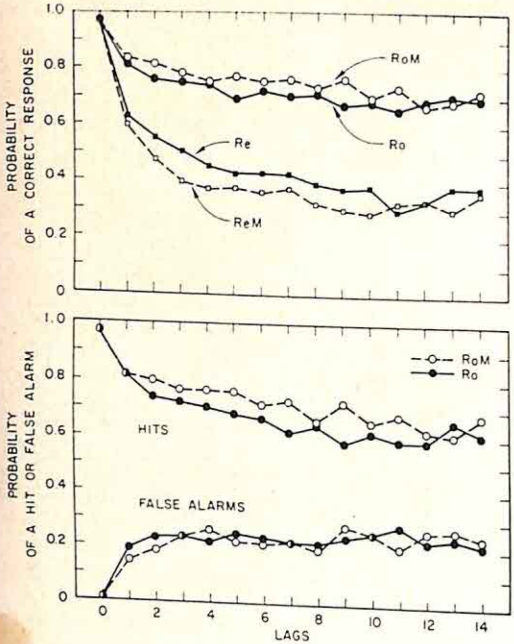

FIG. 1. Top panel—Probability correct as a function of lag for the Re, Ro, ReM, and RoM conditions. (Bottom panel—Hits and false alarms as functions of lag for the Ro and RoM conditions.)

was again tested on a stimulus chosen randomly from the set of nine digits. This procedure continued for 300 trials. At all times, S's task was to remember the *last* response with which each stimulus had been paired. It should be emphasized that the program allowed each S to go at her own pace. After S responded to a test (or after 30 sec.) the program continued, and S's feedback and the new pairing were immediately printed out.

In this type of continuous-task paradigm, the *lag* of an item being tested is defined as the number of test-study trials which have intervened since the item was last studied. The manner of selecting a stimulus to test results in a geometrically decreasing probability of being tested over lags. An item is tested immediately after being studied with probability 1/9, since the stimulus to be tested is chosen randomly out of 9, and, in general, the probability that the lag is equal to i is $(8/9)^i(1/9)$, since the stimulus of interest is not chosen i times (each time with probability 8/9) and then is chosen with probability 1/9.

RESULTS

It has been found in previous studies using a continuous task (e.g., Atkinson, Brelsford, & Shiffrin, 1967) that a slight warm-up takes place at the beginning of a session. For this reason, the data from the first 25 trials of each session were excluded from the data analysis. In addition, the first session of data from each S was excluded, since this was regarded as a practice session during which Ss were adapting to the task and the equipment.

The top panel of Fig. 1 presents the probability of a correct response as a function of lag for the Re, Ro, ReM, and RoM conditions. All curves decreased with increasing lag as expected, and overall performance was quite good. Even with 14 intervening items, performance was far above the chance levels of .04 for recall and .50 for recognition. For purposes of analysis, the recognition conditions are broken into "hits" (the probability of responding "yes" given that the correct response was presented) and "false alarms" (the probability of responding "yes" given that a foil was presented). These probabilities as functions of lag are presented in the bottom panel of Fig. 1. While the hit curves appear to steadily decrease, the false-alarm curves rise to an asymptote at about Lag 2 or 3 and then remain quite stable. For the Re and Ro curves, there are about 1,100 observations at Lag 0. Accordingly, for the ReM, RoM, and the hit and false-alarm curves corresponding to the Ro condition, there are about 550 observations at Lag 0, and for the hit and false-alarm curves corresponding to the RoM condition, there are about 275 observations at Lag 0.

Of crucial importance is the fact that differences exist between the Ro and RoM curves and between the Re and ReM curves. In order to test these differences, the response probability, pooled over all lags, was computed for each S for each curve. To compare any two curves, a t test for matched pairs was performed. For Re versus ReM and for Ro versus RoM, both these tests were significant, $t(7) = 2.35$, $p \approx .05$ and $t(7) = 2.84$, $p < .05$. The two hit curves were found to be significantly different, $t(7) = 2.75$, $p < .05$, but the false-alarm curves did not differ significantly, $t(7) = .65$, $p > .20$. This indicates that the difference between the two recognition conditions is due primarily to the hit rates.

A problem in interpreting these results arises from the fact that the nature of the activity intervening between study and test of a given item differed in the various conditions. Only recognition tests intervened in the recognition condition and only recall tests intervened in the recall condition, while both types of tests intervened in the mixed condition. Conceivably, the differences between the Re and ReM items and between the Ro and RoM items could be explained by assuming that a recognition test produces more interference than a recall test. To investigate this possibility, the following analysis was performed: In the mixed condition, items were examined which had either all recall or all recognition tests intervening between the time the item was studied and the time it was tested. These items were further subdivided according to whether they themselves had been tested by recall or by recognition. For items tested by recognition, the response probabilities were .801 and .824, respectively, for all recall or all recognition tests intervening. For items tested by recall, the corresponding probabilities were .609 and .606. Neither the main effect of intervening activity nor the Intervening Activity × Type of Test interaction approached significance (both Fs < 1). Therefore, the notion that intervening recognition tests generate more interference than intervening recall tests cannot be used to explain the present data.

The main results of this study provide support for the notion that storage differences exist between recognition and recall, contrary to the findings of Freund et al. (1969). In terms of the notation introduced above, the storage operator $\mathbf{S}_{re}$ is apparently more efficient than $\mathbf{S}_m$ which in turn is better than $\mathbf{S}_{ro}$. The latter is somewhat surprising since it might be expected that *S*s would be better at storing information when they are aware of the type of test to be employed. This hypothesis, however, is not supported by the fact that RoM is bettern than Ro.

Thus, in the present experiment, storage differences between recall and recognition appear quite clearly. It is now of interest to make a somewhat more detailed examination of processes occurring at the time of *retrieval*. It has been shown (Freund et al., 1969) that when storage factors are held constant, the relative superiority of recognition versus recall is highly dependent upon the type of procedure used to correct for the disparate guessing rates between the types of test. In the present experiment, application of two such procedures illustrates this dependency. The first, which has been used frequently, is to assume that if S does not know the correct answer, he guesses randomly among the response alternatives (Hilgard, 1951). In this case, let the probability of knowing the correct response be p'. The probability of *not* knowing the correct response and guessing correctly is $(1 - p')\left(\frac{1}{N}\right)$, where N is the number of response alternatives (in this case, 2 for recognition and 26 for recall). The observed probability correct, p, is then

$$p = p' + \frac{(1 - p')}{N}.$$

In terms of our analysis, $p = \mathbf{R}(\mathbf{S}(\mathbf{I}))$, this correction for guessing is now regarded as follows: Whereas it was assumed above that $\mathbf{R}$ encompassed *both* recovery of information from memory and response production, we postulate a new operator, $\mathbf{R}'$, to be only the former process. In other words, $\mathbf{R}'(\mathbf{S}(\mathbf{I}))$ corresponds only to recovered information and can be identified with p'. Now let $\mathbf{G}'$ be the p' to p transformation described by the above equation. Thus,

$$\begin{aligned} p &= \mathbf{G}'(p') \\ &= \mathbf{G}'(\mathbf{R}'(\mathbf{S}(\mathbf{I}))). \end{aligned}$$

Since $\mathbf{G}'$ transforms p' to p, $(\mathbf{G}')^{-1}$ may be applied to observed values of p to obtain

$$\begin{aligned} p' &= (\mathbf{G}')^{-1}(p) \\ &= \mathbf{R}'(\mathbf{S}(\mathbf{I})). \end{aligned}$$

By comparing these p' values ("corrected" probabilities) for the ReM and RoM con-

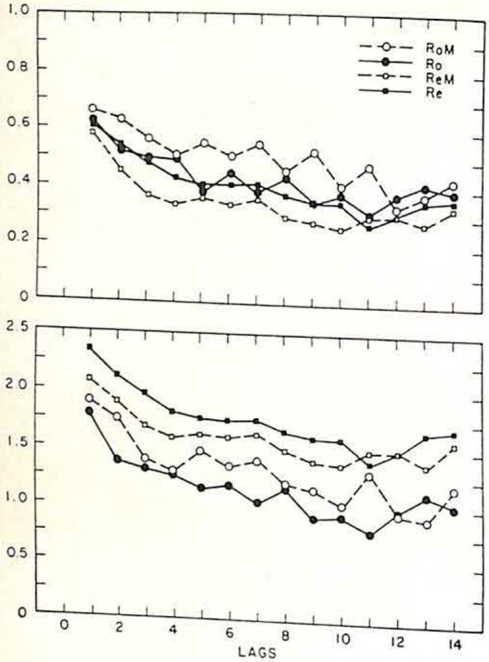

FIG. 2. Values of p' (top panel) and d' (bottom panel) for the Re, Ro, ReM, and RoM conditions.

ditions, we may directly compare our operators, $\mathbf{R}'_{re}$ and $\mathbf{R}'_{ro}$. Since the storage operators, $\mathbf{S}_m$ are the same, in these conditions, differences in p' must be attributable only to differences in $\mathbf{R}'$. The top panel of Fig. 2 shows p' as a function of lag for the Re, Ro, ReM, and RoM conditions. Performance on RoM is superior to that on ReM, $t(7) = 13.0$, $p < .01$. In terms of the above analysis, then, $\mathbf{R}'_{ro}$ leads to better performance than does $\mathbf{R}'_{re}$. The p' curve for Ro was also above that for Re, although not significantly so, $t(7) = 1.16$, $p > .10$.

The second method for comparing recognition and recall is suggested by the theory of signal detectability (TSD) which has been successfully applied to recognition memory (Kintsch, 1968; Murdock, 1965). It is postulated that the strength of information about items in memory may be represented by a single value, d'. Given the probability correct for an N-alternative forced-choice test, or the hit and false-alarm rates for a yes–no test, it is possible to find the corresponding value of d' (Elliott, 1964). Again, in terms of our operator notation, let $d' = \mathbf{R}''(\mathbf{S}(\mathbf{I}))$. Then,

$$p = \mathbf{G}''(d') = \mathbf{G}''(\mathbf{R}''(\mathbf{S}(\mathbf{I}))),$$

where $\mathbf{G}''$ is the transformation of d' to response probability as specified in Elliott (1964). Again, $(\mathbf{G}'')^{-1}$ is explicitly defined and

$$d' = (\mathbf{G}'')^{-1}(p) = \mathbf{R}''(\mathbf{S}(\mathbf{I})).$$

The bottom panel of Fig. 2 presents d' as a function of lag predicted from the data for the Re, Ro, ReM, and RoM conditions. The results are as follows: Re is now superior to Ro, $t(7) = 4.52$, $p < .01$, and ReM is superior to RoM, $t(7) = 4.24$, $p < .01$. Again examining the mixed conditions where the storage operators are the same $\mathbf{R}''_{re}$ leads to better performance than $\mathbf{R}''_{ro}$ which is a reversal of the findings which obtained when $(\mathbf{G}')^{-1}$, the p' transformation, was applied to the data.

This reversal was also obtained by Freund et al. (1969). They applied the p' and d' corrections to their data and found that the former transformation showed recognition to be superior to recall, whereas the latter transformation showed the opposite to be true. Previous studies comparing recall and recognition have, in many cases, used the p' correction for guessing and have concluded that recognition is superior to recall (Postman, 1950; Postman et al. 1948). Studies comparing recall and recognition using a TSD analysis have generally assumed that performance based on information acquired under identical conditions should lead to the same value of d' independent of the type of test. As suggested above, however, it is not meaningful in the present experimental paradigm to make unconditional conclusions regarding the relative superiority of recall versus recognition. We have examined the view that some guessing correction, $\mathbf{G}$, defined for recall and recognition operates on retrieved information to produce a response. In this framework, application of $\mathbf{G}^{-1}$ to the data should produce "corrected" measures which are comparable for the two types of

tests. To reiterate the conclusions of Freund et al., however, the choice of **G** and the resulting comparison of recall and recognition must rest on specific assumptions regarding the nature of the retrieval process.

Discussion

We now turn to an analysis of the data in terms of a theory of memory which has been proposed by Atkinson and Shiffrin (1968). The model to be used in the present article represents an application of this theory to recognition memory and is described in detail by Freund et al. (1969) and Atkinson and Wickens (1971). Basically, the model postulates two memory states: a short-term store (STS) and a long-term store (LTS). Incoming information enters STS where it decays within a short period of time unless it is entered into and is maintained in a rehearsal buffer. Items are assumed to enter this buffer with probability α. The buffer is assumed to have a fixed capacity of r items; *S* may or may not enter a new item into the buffer, but if he does so, a randomly selected item currently in the buffer is knocked out and rapidly decays from STS. Information about an item is transferred to LTS from STS in two ways: (*a*) An amount of information θ' is initially transferred by virtue of the fact that the item entered STS, and (*b*) if the item enters the rehearsal buffer, additional information is transferred at a constant rate θ for each trial that the item resides in the buffer. After the item leaves the buffer, information about it in LTS decreases exponentially at a rate $1 - \tau$ per trial. For purposes of simplicity, θ is set equal to θ'; thus the amount of available information at Lag i about an item which had been in the buffer for j trials ($j \leq i$) is equal to $(j + 1)\theta\tau^{i-j}$. If an item is in the rehearsal buffer when it is tested, a correct response is made with probability one. If it is not in the buffer, a response is made based on information retrieved from LTS. Here, the probability of a correct response is given by a TSD analysis where $d' = (j + 1)\theta\tau^{i-j}$. For yes–no tests, *S* has a response bias for responding "yes." This bias (or criterion), c, is another parameter of the model. For recall, the model thus has four parameters to be estimated: α, r, θ, and τ. For recognition an additional parameter, c, is estimated.

The theory makes a distinction between structural features of memory (e.g., STS, LTS) and control processes used by *S* to deal with specific tasks. Different strategies (control processes) may lead to widely varying values of α, r, and θ. In the present experiment, storage differences may be thought of in terms of these control processes (Atkinson & Wickens, 1971); a different buffer size, for example, might be used by *S* depending on whether he knows he will be tested by recall or by recognition.

TABLE 1

Parameter Values and Chi-Squares for Three Conditions

Parameter	Recognition	Mixed	Recall
α	.79	.73	.53
r	1	2	3
θ	.79	.52	.30
	.95	.97	.99
c	.71	.62	—
χ^2	22.3	29.3	11.3
df	23	37	10

Note.—A dash indicates that the parameter is not needed for the recall task.

A fit of the model to the observed data was made using a minimum χ^2 procedure (Atkinson, Bower, & Crothers, 1964). Three separate parameter estimates were made: one for each of the three study conditions. For the recognition condition, the fit was made over the hit and false-alarm curves derived from the Ro condition, and for the recall condition, the fit was made to the Re curve. For the mixed condition, the parameters were estimated for the recall and the hit and false-alarm data simultaneously. The parameter estimates along with the corresponding χ^2 values and degrees of freedom are shown in Table 1. In terms of the model, the parameter values suggest that *Ss* used quite different control processes depending on the type of test they were anticipating. For the Ro condition, each item has a good deal of information about it transferred to LTS during one trial (the buffer size is 1 and α and θ are fairly high). For Re, on the other hand, there seems to be more emphasis on trying to maintain items in STS without a great deal of effort to store them in LTS (the buffer size is 3 and θ is low). For the mixed condition, as might be expected, an intermediate strategy is indicated (the buffer size is 2, and α and θ have intermediate values). The *Ss'* verbal reports indicated that they were, in

general, using the strategies suggested by the parameter values for the various conditions.

The different control processes for recall and recognition outlined above are logically consistent with the type of information needed for a recall as opposed to a recognition test. For recognition, minimal information about a response is often sufficient to generate a correct response. For example, if *S* can retrieve simply the fact that "the answer rhymes with A," this is enough to be correct if he is presented with "Q" and asked to respond yes or no. A good strategy would thus be to generate as much information as possible about each item and allow the information to decay away since it is still useful in a degraded form. For recall, on the other hand, such degraded information is not as useful, and there would be more reason to try to maintain complete information about as many items as possible in STS where it can be retrieved perfectly.

The strategy for recognition is similar to one suggested by the same model in a previous study. Freund et al. (1969) used a continuous-task paradigm analogous to the mixed condition of the present experiment. There were four types of tests: yes–no, 2, 4, and 26 forced choice, and *S* never knew at the time of study how an item would be tested. The yes–no, 2, and 4 forced-choice tests might well be regarded as recognition tests; thus a reasonable strategy, in terms of the above analysis, would be to store for recognition. When the model described above was applied to the data, the best fit was obtained with a buffer size of 1, α of .75, θ of .86, and τ of .95: These parameter values are remarkably similar to those obtained in the Ro condition of the present study.

There are several conclusions to be drawn from the findings of the present experiment. First, storage differences between recognition and recall have been found contrary to the findings of Freund et al. (1969). As indicated above, the major reason for this is probably that the methodology of the present study is better attuned to the separation of storage strategies. Second, analysis of the data in terms of the memory model discussed above has been highly successful. The model has provided an excellent fit to the observed data, and the obtained parameter values are interpretable in terms of possible storage strategies used by *S* in each of the three study conditions.

REFERENCES

Atkinson, R. C., Bower, G. H., & Crothers, E. J. *An introduction to mathematical learning theory*. New York: John Wiley, 1964.

Atkinson, R. C., Brelsford, J. W., Jr., & Shiffrin, R. M. Multi-process models for memory with applications to a continuous presentation task. *Journal of Mathematical Psychology*, 1967, **4**, 277–300.

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*. Vol. 2. New York: Academic Press, 1968.

Atkinson, R. C., & Wickens, T. D. Human memory and the concept of reinforcement. In R. Glaser (Ed.), *The nature of reinforcement*. Columbus, O.: Merrill Publishing Co., in press.

Elliott, P. B. Tables of d'. In J. A. Swets (Ed.), *Signal detection and recognition by human observers*. New York: John Wiley, 1964.

Freund, R. D., Brelsford, J. W., Jr., & Atkinson, R. C. Recognition vs. recall: Storage or retrieval differences? *Quarterly Journal of Experimental Psychology*, 1969, **21**, 214–224.

Freund, R. D., Loftus, G. R., & Atkinson, R. C. Applications of multi-process models for memory to continuous recognition tasks. *Journal of Mathematical Psychology*, 1969, **6**, 576–594.

Hilgard, E. R. Methods and procedure in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: John Wiley, 1951.

Kintsch, W. An experimental analysis of single stimulus tests and multiple-choice tests of recognition memory. *Journal of Experimental Psychology*, 1968, **76**, 1–6.

MacDougall, R. Recognition and recall. *Journal of Philosophy*, 1904. **1**, 299–333.

Murdock, B. B. Signal detection theory and short-term memory. *Journal of Experimental Psychology*, 1965, **70**, 443–447.

Postman, L. Choice behavior and the process of recognition. *American Journal of Psychology*, 1950, **63**, 576–583.

Postman, L., Jenkins, W. O., & Postman, D. L. An experimental comparison of active recognition and recall. *American Journal of Psychology*, 1948, **61**, 511–519.

(Received March 22, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 227–232

AMOUNT OF INFORMATION AND INTRALIST SIMILARITY IN PAIRED–ASSOCIATES LEARNING

OVID J. L. TZENG[1]

Pennsylvania State University

Two experiments were performed to assess the roles of amount of information and intralist similarity in paired-associates (PA) learning. By definition, amount of information contained in a PA list is an inverse function of intralist similarity in the same list. Their effects on learning were found to be opposite in direction, depending upon what kind of material was used. Greeno's two-stage (learning to store vs. learning to retrieve) memory model for PA learning was employed to account for the data obtained. It was hypothesized that amount of information was critical only at the storage stage and intralist similarity was critical only at the latter stage. The results seemed to confirm this hypothesis. The present study suggests an experimental way to get at the major locus of shifts in parameter values.

When amount of information is defined as the amount by which uncertainty has been reduced, the major theoretical concept is the probability of occurrence of each alternative. Learning, in a sense, can be viewed as the case that after a period of training, *S* is able to retrieve what has been stored and to make an appropriate response. Specifically, *S* produces his response from an information space. When this space can be specified, there is no doubt that the amount of information involved is an important variable. Riley (1952), in a paired-associates (PA) experiment, showed that the greater the response uncertainty, the slower the learning. Brogden and Schmidt (1954) carried out an analogous experiment using a verbal maze with variation of response routes in each response unit. The same effect was observed. Adelson, Muckler, and Williams (1955), in a series of experiments, concluded that uncertainty is an operative factor in verbal learning. One apparent contradiction arises in the case of a list of trigrams constructed from a small number of letters or symbols. According to the information theorists, the learning of such a list should be rapid. On the other hand, the interference theorist would suggest that the smaller number of letters used to construct the list, the higher the formal intralist similarity with the consequence of slower acquisition. This latter phenomenon was observed by Gibson (1942) and Underwood (1956).

The effects of amount of information and of intralist similarity on verbal learning are, then, found to be opposite in directions. But by definition, the amount of information is an inverse function of intralist similarity of the same list. The important problem is to identify the operating variable, given any verbal list. A resolution can be reached by applying Greeno's (1970) two-stage memory model of PA learning. The first stage is mainly a stage of encoding and storing the stimulus and response as a whole unit. The amount of information involved is an effective variable at this stage. When each item has been coded into its representational form, learning proceeds to the retrieval stage at which the essential process is to eliminate the confusion among the items and to retrieve reliably the correct items. The effect of amount of information disappears, while intralist similarity becomes operative at this stage.

If the above interpretation is correct, it should be possible to produce the same patterns resulting from the effect of either the amount of information or intralist

[1] The author is grateful to Charles N. Cofer and Francis J. DiVesta for their critical reading and suggestive comments on this paper. The assistance of Joyce Birmingham and Carl Harris in correcting some grammatical errors of the writing is acknowledged with appreciation.

Requests for reprints should be sent to Ovid Tzeng, Department of Educational Psychology, 201 Social Sciences Building, Pennsylvania State University, University Park, Pennsylvania 16802.

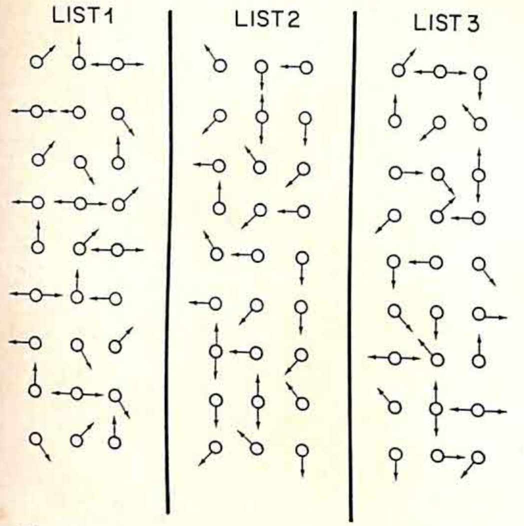

FIG. 1. Three experimental lists for Exp. I.

similarity. With verbal trigrams being used, the amount of information involved in the first stage of learning should be relatively unimportant due to the fact that the higher order habits and skills required for the integration of verbal units have already been in *S*'s long-term or permanent memory. However, the verbal properties such as acoustic or articulatory confusion should enhance the similarity effects. On the other hand, when the materials are nonverbal figures which are so unfamiliar that the components of items themselves and their integration must be "learned," the amount of information becomes operative at the first stage, and their nonverbal characteristics should minimize the similarity effects at the second stage.

The present experiments attempted to manipulate the degree of difficulty demanded at each stage by means of the combination of two specified information spaces (i.e., 5 vs. 10) and two different kinds of material (i.e., CCC vs. nonverbal figure set). Furthermore, a special technique (the fragment-cue method; see Liu, 1969) was employed as the measure of acquisition in order to examine the internal formation of each item during acquisition. The three sizes of fragment cue (i.e., 2, 1, and 0) then represent three special cases of constant order paired associates with the response terms being defined as letters, bigrams, and trigrams, respectively. (Corresponding response terms were arranged for the nonverbal figures.)

METHOD

Materials.—The two experiments were virtually identical except that one used nonverbal figures to construct the list, and one used CCC trigrams. The lists all consisted of nine items. For the nonverbal figure experiment, the 10 nonverbal figures were ô, ♂, ↖o, ←o, o→, ♀, o↘, ↙o, ô̦, and ←o→. Five of them (ô, ♂, o↘, ←o, and ←o→) were used to construct List 1, and the other 5 were used for List 2. List 3 contained all 10 figures (Fig. 1).

For the CCC experiment, the list was constructed from either 5 or 10 consonants. The list made from 5 letters was X S F, S N C, C S X, N F S, X C F, C N X, S X F, F C S, and N F C. The list made from 10 letters was N T P, P D X, C S W, B F D, D W S, F X C, X C N, W B T, and T N F. The associative value of each item in the 5-letter list was equated with the associative value of one of the trigrams in the 10-letter list. This one-to-one correspondence made these two lists have the same amount of associative value (39.78% on the average; see Witmer, 1935).

For all the lists, each alternative had an equal probability of occurrence within each list, with the restriction that no repetition of the same letter was allowed within each item.

Subjects.—The *S*s were 52 undergraduates enrolled in an educational psychology course at Pennsylvania State University. They had never served as *S*s in any psychological experiment. They were given additional credits for participation. While the nonverbal figure experiment (30 *S*s) was run in its entirety before the CCC experiment, within each experiment *S*s were assigned to groups in order of appearance at the laboratory according to a scheme randomizing the order of each condition.

Procedure.—First, *S* was given a practice task. Three items (ô ←o ♀, o→ ô ←o, and ô ♀ ←o for the nonverbal figure experiment and A B C, D E F, and P Q R for the CCC experiment) appeared one at a time on a memory drum (C. H. Stoelting Co.). Each item lasted 2 sec. and a 2-sec. blank space separated one item from the next. After *S* studied this short list, a fragment of one set of the figure (or trigram) of the list appeared, and *S* was asked to write down the missing part. The *S* was tested with the fragment ô ←o __, then __ __ __, and then ô __ __ (correspondingly A B __, then __ __ __, and then P __ __ for CCC). In responding to the second "fragment," *S* was reminded to study not only the item but also its position in the list.

After *S* completed the practice task, the experimental list began to appear on the memory drum. The nine items appeared one at a time. Each item lasted 2 sec., and a 2-sec. blank space separated one item from the next. The list was presented in the same order throughout. After every four presentations of the list, a test trial was administered. On

TABLE 1

RESULTS OF EXPERIMENT I: PROPORTION OF SUBJECTS GIVING CORRECT RECALL TO THE THREE SIZES OF FRAGMENT CUE FOR EACH TREATMENT GROUP ON THE FIVE TRAINING BLOCKS

Information space	0					1					2				
	Test trial					Test trial					Test trial				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
a. Recall of the whole															
5	.00	.18	.12	.22	.18	.11	.11	.24	.36	.55	.30	.43	.68	.64	.73
10	.00	.05	.05	.05	.15	.04	.08	.12	.18	.28	.14	.26	.34	.57	.54
b. Recall of the last missing figure															
5	.09	.30	.41	.37	.30	.30	.33	.36	.45	.62	.30	.43	.68	.64	.73
10	.05	.09	.12	.12	.27	.04	.10	.21	.33	.30	.14	.26	.34	.57	.54

a test trial, a fragment of each item of the list was presented in the serial order of the items in the list. The *S* was allowed as much as 4 sec. for starting to write down the missing parts. If he did not start to respond within 4 sec. after the presentation of a fragment, the fragment of the next item appeared. When a fragment was presented, it was always the first or the first two parts. Some tests involved no fragment at all. On each test trial, each item was tested by only one kind of fragment, each kind of fragment appearing three times in any one test trial. An example of a test for the figure list is as follows: ♂ — —, — — —, — — —, ⟲ — —, ô ♂ —, ⟲ ô —, ⟲ ⚲ —, ô — —, and — — —.

For each *S*, the acquisition list was presented 20 times. Hence, five test trials were adminstered for each *S*. On each test trial, a different test list was used. The order of assignment of test lists was randomly varied from one test trial to another within *S* and from *S* to *S*.

RESULTS

Figure experiment.—Table 1 shows the proportion of *S*s giving correct recall to each type of stimulus fragment under two scoring procedures (1a, 1b). A .05 level was employed as statistical significance criterion.

1a. Recall of the whole: The first scoring demanded that every missing part be filled with the correct figure. Table 1a shows the results of this scoring. Obviously, both increases in the size of fragment and in training facilitated recall. The effects of fragment size and training were significant, $F\ (2, 36) = 75.54$, and $F\ (4, 72) = 20.73$, respectively. The learning of the list of 5 figures was significantly better than that of the 10-figure list, $F\ (1, 18) = 8.72$. The only significant interaction was observed for Fragment Size $\times$ Training, $F\ (8, 144) = 2.68$. This effect reflected the unequal responding difficulty required for the three sizes of fragment.

1b. Recall of the last missing figure: Recall was also scored correct if *S* reproduced only the last missing figure correctly. These data are summarized in Table 1b. This scoring procedure represents an attempt to equate the responding difficulty for the three fragment sizes. An analysis of variance indicated similar results as before. The *F* values for the effects of amount of information, fragment size, and training were: $F\ (1, 18) = 15.40$; $F\ (2, 36) = 28.15$; and $F\ (4, 72) = 21.55$, respectively. The relatively large *F* value obtained here in comparison with the first scoring procedure indicates that the second scoring is more sensitive to the effect of amount of information.

CCC experiment.—The data of this experiment are shown in Table 2 under two scoring procedures (2a, 2b).

2a. Recall of the whole: As before, the recall differences were again found as a function of the size of fragment, $F\ (2, 40) = 51.75$, and training, $F\ (4, 80) = 30.75$. However, the amount of information had

TABLE 2

RESULTS OF EXPERIMENT II: PROPORTION OF SUBJECTS GIVING CORRECT RECALL TO THE THREE SIZES OF FRAGMENT CUE FOR EACH TREATMENT GROUP ON THE FIVE TRAINING BLOCKS

Information space	0					1					2				
	Test trial					Test trial					Test trial				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
a. Recall of the whole															
5	.03	.22	.25	.57	.57	.24	.36	.36	.43	.41	.45	.57	.60	.68	.82
10	.15	.40	.60	.65	.82	.23	.52	.63	.77	.80	.66	.72	.84	.96	.99
b. Recall of the last missing letter															
5	.10	.30	.43	.66	.68	.46	.52	.44	.64	.52	.45	.57	.60	.68	.82
10	.18	.42	.68	.68	.82	.22	.52	.72	.77	.80	.66	.72	.84	.96	.99

an opposite effect on recall in comparison with the figure experiment. That is, S learned the 10-letter list significantly faster than the 5-letter list, F (1, 20) = 16.01. So far as intralist similarity is concerned, this result is the same as that obtained in the study by Gibson (1942).

2b. Recall of the last missing letter: When the response probability was equated by using the reproduction of the last missing letter as the score (Table 2b), the three variables, amount of information, the size of cueing fragment, and acquisition trials, all had the same effects as shown in the first scoring procedure (Table 2a). Their F values were: F (1, 20) = 15.93; F (2, 40) = 35.16; F (4, 80) = 30.51, respectively. It should be noted that the only significant interaction was between training trial and amount of information (or properly, intralist similarity). An inspection of Fig. 2 shows that as the training goes on, intralist similarity becomes more and more effective in the later stage of learning.

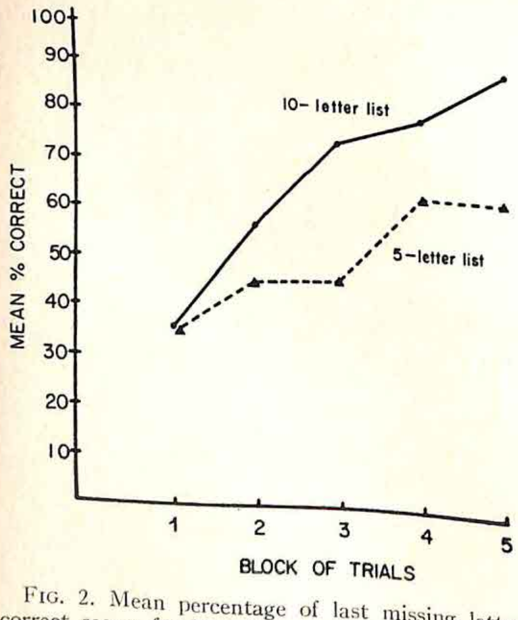

FIG. 2. Mean percentage of last missing letter correct scores for both 5-letter and 10-letter lists as a function of training blocks.

DISCUSSION

Garner (1962) and many others have reported data indicating that the amount of information transmitted in recall increases directly with the formal uncertainty of the ensembles of items used as memory materials and impairs acquisition. The present study shows the insufficiency of this interpretation when applied to some verbal lists where intralist similarity is an effective variable. As indicated above, the amount of information contained in a list, whether verbal or nonverbal, is an inverse function of the formal intralist similarity in the same list. The important thing is to determine in what situations learning will be influenced by one factor and not by the other. Greeno's (1970) two-stage model

of PA learning seems to be able to give a clear interpretation for these opposite phenomena. By varying the degree of difficulty demanded at either the storage or the retrieval stage, the present study produced two different patterns of results, each proposed by either the information theorists or the interference theorists. The present study suggests an experimental way to get at the major locus of shifts in parameter values.

The notion that the dependence of the storage stage on, and the independence of the retrieval stage of, the amount of information is powerfully supported by a recent study by Miller and Weinstock (1971). Predefined high and low memory span (H-MS, L-MS) *S*s learned a PA list with responses either high or low in formal intralist similarity. It was found that the L-MS *S*s made more errors before the first correct anticipation for each pair than did the H-MS *S*s. However, such differences due to different level of MS disappeared afterwards. On the other hand, the notion that the intralist similarity is only critical at the later stage of learning is evidenced by the study of Runquist (1966). It was found that learning was affected by intralist similarity only when the errors after the first correct were considered (cf. Greeno, 1970).

The above reasoning has another implication. If an item has been learned at the first stage but not yet at the second stage, it should be expected that recognition of the item is possible. However, reliable recall would still require the completion of the second stage. When verbal material is used, the first stage of learning should be released relatively early from the effect of amount of information, but the second stage of learning should be subject to the effect of intralist similarity. That is, if the intralist similarity is varied between lists, the model predicts that recall should be suppressed in the high intralist similarity condition, but the recognition of the items in both conditions (i.e., high vs. low intralist similarity) should show no difference. Kintsch (1968) reported an experiment that confirmed this prediction.

A recent concern in the literature has been with examining the relationship of similarity to meaningfulness. Underwood and Richardson (1956), using four CVC lists with two levels of Glaze association value (93.3% to 100% vs. 0% to 20%) and two levels of intralist similarity, demonstrated that learning was an inverse function of the degree of intralist similarity; the difference in learning between high and low intralist similarity was greater for lists containing the low association syllables. Despite the procedural differences, the present CCC experiment confirms their results. However, Underwood interpreted this phenomenon in terms of associative (or hook-up) learning, while the present model predicted its occurrence in terms of the retrieval process. When meaningfulness is increased, the formal intralist similarity becomes less effective due to the fact that the retrieval cues can be coded in terms of either meaning or their acoustic character in order to overcome the interference. The present model emphasizes the active role of *S* in searching for a reliable retrieval strategy.

One further point should be mentioned. It is concerned with the internal structure of the items in a list (Garner, 1962) or the identity of letter location (Runquist, 1971). From the viewpoint of the present model, one must be cautious in interpreting the data in terms of intralist similarity only. For when some of the letters consistently occur in a certain position, the amount of information is reduced. And this fact should be taken into account for it does affect the earlier stage of learning. Take the Neutral condition of Runquist's (1971) experiment, for example. If it is assumed that his practice list represents the earlier stage of learning and his test list represents the later stage of learning, it is clear that in the practice list, the order of mean correct went down from HS-1 to HS-2, and then to LS although the differences were relatively small. However, the order for the test lists went up from HS-2 to HS-1 and then to LS, and the effect was significant. So far as the shift of the parameter from amount of information to intralist similarity as learning proceeds from storage process to retrieval process is concerned, the present model is able to account for these facts.

REFERENCES

ADELSON, M., MUCKLER, F. A., & WILLIAMS, A. C., JR. Verbal learning and message variables related to amount of information. In H. Quastler (Ed.), *Information theory in psychology*. Glencoe, Ill.: Free Press, 1955.

BROGDEN, W. J., & SCHMIDT, R. E. Acquisition on a 24-unit verbal maze as a function of number of alternate choices per unit. *Journal of Experimental Psychology*, 1954, **47**, 335–338.

GARNER, W. R. *Uncertainty and structures as psychological concepts*. New York: Wiley, 1962.

Gibson, E. J. Intra-list generalization as a factor in verbal learning. *Journal of Experimental Psychology*, 1942, **30**, 185–200.

Greeno, J. G. How associations are memorized. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

Kintsch, W. Recognition and free recall of organized lists. *Journal of Experimental Psychology*, 1968, **78**, 481–487.

Liu, I. M. Cueing function of fragments of verbal items. *Journal of Experimental Psychology*, 1969, **82**, 107–114.

Miller, S. M., & Weinstock, R. B. Memory span and intralist response similarity in paired-associate learning. *Psychonomic Science*, 1971, **23**, 139–140.

Riley, P. A. Rote learning as a function of distribution of practice and the complexity of the situation. *Journal of Experimental Psychology*, 1952, **43**, 88–95.

Runquist, W. N. Intra-list interference as a function of list length and inter-stimulus similarity. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 5–13.

Runquist, W. N. Stimulus coding and interference in paired-associate learning. *Journal of Experimental Psychology*, 1971, **87**, 373–377.

Underwood, B. J., & Richardson, J. The influence of meaningfulness, intralist similarity, and serial position on retention. *Journal of Experimental Psychology*, 1956, **52**, 119–126.

Witmer, L. R. The association value of three-place consonant syllables. *Journal of Experimental Psychology*, 1935, **47**, 337–360.

(Received April 10, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 233–239

INFLUENCE OF CONCURRENT AND TERMINAL EXPOSURE CONDITIONS ON THE NATURE OF PERCEPTUAL ADAPTATION [1]

JOHN J. UHLARIK [2] AND LANCE KIRKPATRICK CANON [3]

University of Washington

The *S*s viewed their own localizing movements through a laterally displacing prism as they pointed at a visible target. In the Concurrent Exposure condition, the pointing arm was visible throughout its excursion from resting place to target, while in the Terminal Display condition it could be seen only at the termination of a pointing movement. L. K. Canon's model of the process of adaptation holds that compensatory shifts in localization manifest themselves primarily in the modality not attended to or employed as a source of information for localizing responses. With terminal display conditions during exposure to the intermodality inconsistency, where *S*s were likely to attend to proprioceptive cues in making their localizations, subsequent shifts in the position selected as the visually straight ahead were found. With concurrent display conditions, where exposure period localizations would be expected to be based on visual cues, shifts in the arm position felt to be straight ahead occurred. The relevance of these findings to prior research on interlimb transfer of adaptation was discussed.

That individuals will compensate for the errors of localization induced by various rearrangements of the inputs to sensory receptors is amply documented. Several attempts have been made to provide general statements adequate to account for the observed conditions under which such adaptation develops and to delineate the nature of the process involved (e.g., Harris, 1965; Held, 1962). However, none of these attempts has proved entirely satisfactory in the light of subsequent research. Recently Canon (1966; 1970; 1971) has proposed an alternative formulation which holds that adaptation will develop as the result of intermodality inconsistency of inputs. Briefly, the model suggests that when an individual experiences inconsistent or conflicting information from two or more modalities relevant to the spatial location of some external stimulus object, an adjustment to these conditions will develop. Further, it is predicted that the resultant adaptation or recalibration will manifest itself in the modality not utilized as a source of information on which the response to such a stimulus object was based. The end result of this process will be a state of affairs in which *S* will, when required to localize some object, select the same spatial position no matter which modality is used to guide the localizing responses. In other words, adaptation involves recalibration of these inputs so that the intermodality conflict is functionally eliminated and once again the inputs yield redundant or complementary information about spatial location.

Previous research designed to investigate the adequacy of this formulation has directly manipulated the attention variable by means of explicit instructions to *S*s regarding the modality to which they were to attend during exposure. However, its applicability is not limited to this experimental paradigm. The physical conditions of exposure to intermodality inconsistency might well play a role in producing differences in the modality to which an *S* will attend during the exposure period. Howard (1968), from a different analytical stance, has suggested a categorization of training or exposure circumstances employed in prism-adaptation research where exposure

[1] This research was supported by Grant GB-7693 from the National Science Foundation to the second author. We wish to thank Patrick Sweeney and Dennis Healey for their assistance in conducting the experiment and Robert Abbott for his helpful comments on the manuscript.

[2] Now at Kansas State University.

[3] Requests for reprints should be sent to Lance K. Canon, University of Washington, Department of Psychology, Seattle, Washington 98105.

to the sensory discordance is provided by having *S*s view their arms while localizing visual targets. One of these training procedures, termed "terminal display," refers to conditions in which "the subject is allowed to see his hand only at the termination of an aiming movement [p. 25]" and the other, referred to as "concurrent display," involves circumstances in which "the subject views his hand continuously while aiming at a target [p. 20]." With reference to the concurrent display, he goes on to suggest that "if the subject is allowed to view his hand while he is moving it, he may visually guide it onto the target and ignore the discordant proprioceptive and motor activity [p. 25]." Howard's analysis of the consequences of these types of training procedures leads him to the conclusion that terminal display conditions will lead to greater or more complete adaptation.

Within the frame of reference of the present model, this distinction leads to predictions not about the relative magnitude of the consequent adaptation but rather concerning the specific type of adaptation that will develop. Concurrent display conditions would be expected to involve, as Howard (1968) notes, circumstances in which *S*'s tendency will be to attend to the visual information when localizing targets during exposure. Research on visual dominance or swamping (e.g., Gibson, 1933; Hay, Pick, & Ikeda, 1965; Rock, 1966) suggests that *S*s will spontaneously depend upon visual cues in guiding their localizing movements under conditions in which they are permitted to view their pointing arm continuously. Terminal display conditions, on the other hand, are likely to force *S* to attend to proprioceptive cues and to depend upon them in guiding his hand to the target since he is permitted to view his finger only at the terminus of its excursion toward the target.

To provide evidence relevant to the predictions the model makes regarding the specific nature of the adaptation developing under these two types of exposure conditions, the dependent measures of pre-exposure to postexposure shifts in localizations were of three types. One permitted testing for visual-motor shifts by having *S* select a visual position that appeared to him to be straight ahead with reference to his midsaggital plane. Another assessed possible proprioceptive-motor adjustments by having *S* point with his arm to what he felt to be the straight-ahead position under conditions not permitting any visual feedback from his limb. No visual target was involved in this measure as *S*'s eyes were closed. The final measure was typical of those taken in most studies of adaptation, as *S* pointed at a visible target without visual feedback from his limbs. Such a measure would be expected to confound and reflect the independent visual or proprioceptive shifts that may have developed.

Specifically, then, the present model predicts that adaptation developing with concurrent display conditions during exposure will manifest itself in measures sensitive to proprioceptive shifts. Conversely, terminal display exposure circumstances should produce adaptation that will be exhibited through shifts in the measure of visual egocentric localization. No specific predictions regarding the relative magnitude of these effects follow from the model, and thus the only expectation regarding the final measure, which confounds the visual and proprioceptive shifts, is that significant adaptation will be found with either display condition.

METHOD

Subjects.—Sixteen male college undergraduates served as *S*s in the experiment. Participation in experimental research was a requirement for the introductory psychology course in which they were enrolled. Right-hand dominance was a prerequisite and two individuals were rejected because of a history of astigmatism.

Apparatus.—A schematic diagram of the apparatus is given in Fig. 1. The visual display consisted of a row of 61 small lights, 2 mm. in diameter and 1 cm. apart on center, situated 14 cm. below eye level on a vertical panel perpendicular to *S*'s line of sight. Thirty lights were located on each side of the center light which was positioned slightly to the right of *S*'s midsagittal plane in the center of the field of view of *S*'s right eye. The panel circumscribed an arc with a radius of 57 cm. A second set of targets, consisting simply of three stenciled dots 5 cm. apart on center and labeled A, B, and C was

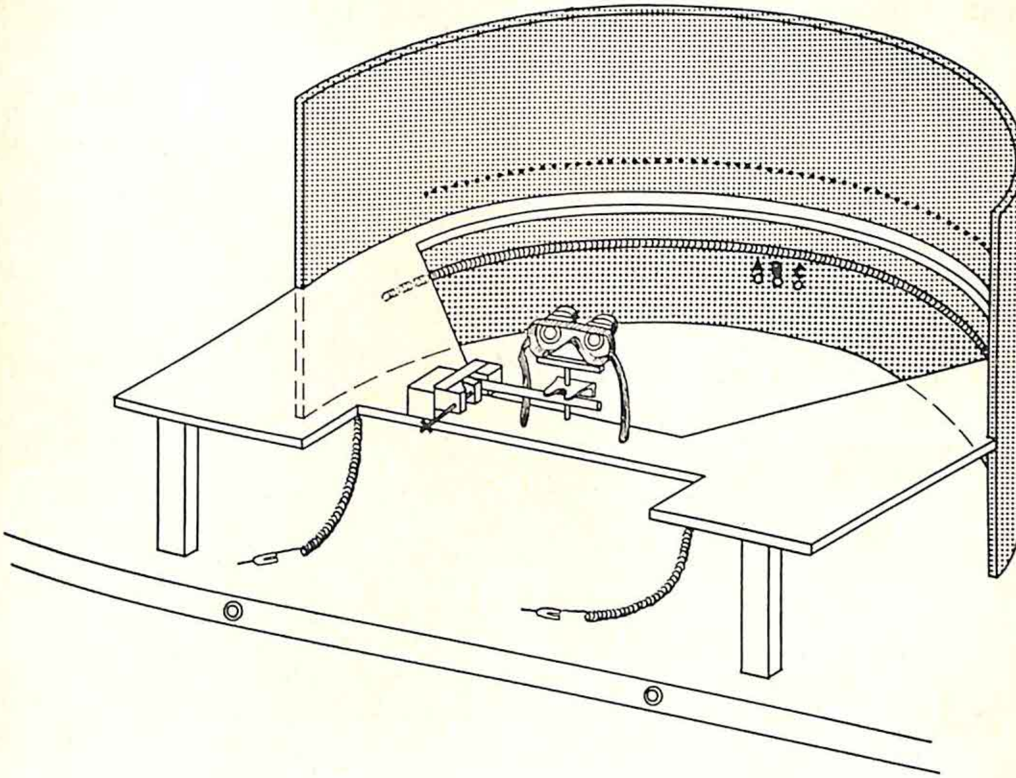

FIG. 1. Prism, bitebar, and localizing apparatus.

located 12 cm. below the row of lights. The dot labeled B was directly beneath the center light.

The *S* viewed the stimulus display from a distance of 57 cm., with his head held in position by a bite-board. When making the dental impression, *S* was instructed to position himself such that the dot labeled B appeared to be straight ahead of his right eye.

Directly beneath the row of stimulus dots was a horizontal Plexiglas rod wound, eight turns per centimeter, with resistance wire. The *S* wore a plastic finger splint with a 2.5-cm. long brass stylus protruding from the end of the index finger of his right hand. Pointing localizations were made by touching the 120-cm. long resistance coil, circumscribing the same arc as the visual display. This was positioned 10 cm. beneath the row of stimulus lights. The *E*, seated behind *S*, could determine the position of *S*'s localizations by reading a digital voltmeter that registered the amount of resistance in a circuit containing the meter, the resistance coil, and a power supply. The *S* completed the circuit by touching the stylus to the coil. A digital readout voltmeter permitted accurate measurement to within .6 mv. (.2°) and was not subject to potential bias or parallax inherent in reading dial-calibrated panel voltmeters.

A black, cloth-covered panel could be placed over the open horizontal surface of the apparatus to prevent *S* from viewing his limbs while measurement of pointing localizations were obtained. All of the other surfaces of the apparatus were covered with black felt cloth or painted flat black.

Visual displacement to the left or to the right, for the right eye, was produced by a Risley rotating prism (Bausch & Lomb No. 71 48-59). The prism was adjusted to a strength of 20 diopters, which should laterally displace the visual field 11.3°. The prism was mounted in the right eyepiece of a pair of welder's goggles and the left eyepiece was occluded throughout the study. The width of the monocular field of view for the right eye was 35°. The frame containing the goggles was rigidly attached to the table at which *S* sat.

Procedure.—In order to familiarize *S* with the apparatus and the manner in which he was to make his localizations, 15 practice trials were given during which the monocular prism goggles set at 0 diopters were worn. The room lights were on and the opaque panel was removed so that *S* could view his arm as he pointed at visible target positions by touching the coil at the appropriate position with the stylus attached to his right index finger. No data were recorded during these trials.

Next, a series of preexposure measures of three different types was made with the prism set at 20 diopters. All localizations were made with the right arm. Contact with the coil was made by reaching under the shoulder-level black panel completely covering the horizontal surface of the apparatus.

The S's head was fixed in the straight-ahead position by the biteboard and no visual feedback from any part of his body was possible. The room was darkened completely except for a very dim light on E's control panel behind S.

Three different types of measurements were taken. One, termed the visual motor (VM), involved having S indicate the position he saw as "straight ahead." He was instructed to use his median or midsaggital plane as the referent here, i.e., to indicate what point appeared to him to be directly in line with the midline of his head and body, or in other words, straight ahead of the tip of his nose. A single light in the horizontal stimulus array was illuminated in the periphery of S's monocular field of view, and S indicated the direction to move the light by pressing a left or a right signal button. The E would then turn off the first light and illuminate the next adjacent one 1° in the direction specified by S. There was no time limit on S's responding and he was instructed to repeat this procedure, including as many reversals in direction as he felt were necessary, until the stimulus light that was illuminated appeared to be straight ahead. This measure was repeated four times, starting at alternate sides in the periphery of the monocular field.

Another measure, the proprioceptive motor (PM), was the arm position S selected as feeling straight ahead. It was obtained by having S point straight ahead with no visual target and no visual feedback regarding his pointing limb. The S was instructed to close his eyes and point "straight ahead," touching the appropriate place on the coil with the stylus. The E recorded the reading on the meter and told S to return his hand to the surface of the table. This measure was also repeated four times.

The final measure required S to point at a series of visible targets, again with no visual feedback from the pointing limb. Since such a gross measure would be expected to reflect both visual and proprioceptive effects, if any were present, it was labeled the visual-proprioceptive-motor index (VPM). The E illuminated the appropriate target light, S made his localization of it, then signaled when he felt he was on target, and returned his hand to the table surface. Three target positions, 0°, 6° left, and 6° right, were employed and each was presented three times in a randomly determined order.

The order of presentation of the VM and PM measures of egocentric localization was counterbalanced. For half of the Ss the VM measures were always taken first, and the remaining half was given the PM measures first. The VPM measures were always obtained last.

Following these preexposure measures, the horizontal panel preventing S from viewing his localizing movements was removed and two 15-w. lights located under the top surface of the apparatus were illuminated so that monocular visual feedback with the 20-diopter prism in place would be possible during the exposure period. All Ss made a total of 60 localizations of the three stimulus dots labeled A, B, and C. The S was to reach out and point to the dot corresponding to the letter called out on a prerecorded tape and then return his hand to a designated starting position on the table. He was instructed not to make corrections should he miss the target on his initial attempt. A letter was called out every 3 sec., and each of the three targets was designated an equal number of times in a randomly determined order. A 20-sec. rest period was given after 30 localizations had been made.

There were two conditions of exposure employed. For Ss who participated in the Concurrent Display condition (half of the Ss), unrestricted observation of their localizing activity was possible. The S's hand was visible to him during the entire excursion from the resting position to the target, a distance of approximately 38 cm. The remaining Ss operated under the Terminal Display condition. Here an opaque panel was installed permitting S to view only the stylus attached to his finger when his arm was extended and the stylus was touching the rear vertical panel of the apparatus in the area of the exposure targets. Thus S could receive visual feedback regarding his localizing movement only after it had been initiated and was near the terminus of the excursion toward the target.

Following the exposure period, the apparatus lights were turned off, the full panel was reinstalled to once again completely eliminate visual feedback, and a set of postexposure measures was taken. The same order and presentation conditions as had been employed during the preexposure measures obtained with S still wearing the 20-diopter monocular prism and head movements restricted by the biteboard.

Results

The primary data of the study came from preexposure to postexposure difference scores on the VM, PM, and VPM measures. For all analyses these difference scores were signed such that positive entries indicated that the represented shift in mean localization was in a direction compensating for prism-induced error.

Table 1 presents the overall mean adaptation scores obtained in the Terminal and Concurrent display conditions as reflected in the three different measures taken in each condition. A further breakdown on the basis of which one of the measurements, VM or PM, was taken first in the test sequence given each S is also given. This order of presentation of the measures variable was counterbalanced in the design of the study and thus the overall adaptation scores, which are of central concern, should not be systematically influenced by any order effects resulting from this repeated measures design. In the Concurrent Dis-

play condition the magnitude of the overall mean VM shift was not significant while the PM shift was. In the Terminal Display condition the opposite relationship obtains—there is a highly significant overall VM shift and a smaller PM shift of borderline significance, $t = 1.59$; $p < .06$. Of primary interest is the predicted interaction, between the conditions of exposure and the measures taken. Table 2 gives the results of a repeated measures, randomized-blocks-design analysis of variance of the adaptation scores given in Table 1. As indicated in this table the Exposure Conditions (A) × Measures (D) interaction is statistically significant, as predicted. To give a more detailed analysis of this interaction, the measures effect was partitioned into relevant orthogonal comparisons (cf. Edwards, 1968). The breakdown comparing the VM and PM measures for the two exposure conditions was highly significant, F (1, 16) = 11.79, $p < .005$, and accounts for 98% of the variation in the A × D interaction. This indicates that the significant variation represented in the A × D interaction was due to variation in differences in the magnitudes of the VM and PM measures for the terminal and concurrent exposure conditions as was expected.

In addition to the variables of direct interest, exposure condition and measures, the effects of the counterbalancing procedures for prism orientation and for order of presentation of measures employed in the study were included in the analysis to permit the extraneous error variance associated with them to be accounted for explicitly. The significant prism orientation (B) effect reflects the relatively smaller mean shift found when the prisms were set to produce displacement to the left ($\bar{X} = 3.1°$) than when rightward displacement was induced ($\bar{X} = 5.1°$). The significant main effect of order of presentation of measures was produced by the smaller overall mean shift when the PM measures were taken first ($\bar{X} = 3.3°$) than was the case when the VM measures preceded the PM measures ($\bar{X} = 4.8°$). The significant D main effect stems from the differences in the magnitude of the VPM shift ($\bar{X} = 5.5°$) relative to those of the VM ($\bar{X} = 3.3°$) and the PM ($\bar{X} = 3.3°$).

TABLE 1

Adaptation Scores in Degrees (Mean Differences Between Preexposure and Postexposure Mean Scores)

Exposure cond.	Measurement cond.	VM	PM	VPM
Concurrent display	VM first	5.3 (±4.4)	1.5 (±1.7)	6.1 (±2.1)
	PM first	−0.8 (±2.5)	6.8 (±2.3)	4.6 (±3.6)
	Overall	2.3 (±3.2)	4.2* (±1.9)	5.3* (±1.6)
Terminal display	VM first	6.8 (±2.9)	2.9 (±4.6)	5.7 (±3.8)
	PM first	2.1 (±2.5)	1.7 (±4.4)	5.8 (±3.4)
	Overall	4.4* (±2.5)	2.3 (±2.3)	5.7* (±1.8)

Note.—Numbers in parentheses are 95% confidence intervals.
* $p < .005$.

TABLE 2

Analysis of Variance of Mean Adaptation Scores

Source of variation	df	MS	F
A Exposure cond.	1	.13	< 1
B Prism orientation	1	45.05	11.88**
C Order of presentation of measures	1	26.85	7.08*
A × B	1	3.15	< 1
A × C	1	2.13	< 1
B × C	1	.40	< 1
A × B × C	1	1.90	< 1
Error (a), S (ABC)	8	3.79	
D measures	2	25.81	8.04*
A × D	2	19.03	5.93*
B × D	2	2.60	< 1
C × D	2	51.37	16.01***
A × B × D	2	1.03	< 1
A × C × D	2	19.08	5.94*
B × C × D	2	1.58	< 1
A × B × D × D	2	4.49	1.40
Error (b), S (ABC)D	16	3.21	

Note.—The use of a randomized blocks design involves the assumptions of homogeneity of variance and equal correlation between treatments. If these assumptions are violated, F is not distributed with $k - 1$ and $(n - 1)(k - 1)$ degrees of freedom. These assumptions were not always met in the present analysis of variance. Thus, a Conservative F test, with 1 and $n - 1$ degrees of freedom was used for all tests of significance involving error b (cf. Edwards, 1968).
* $p < .05$
** $p < .01$
*** $p < .005$

Discussion

In general, the data provide consistent support for the hypotheses advanced. As expected, there is a significant interaction between

the conditions of exposure, terminal or concurrent, and the nature of the adaptation that develops. With Concurrent Display during exposure, *S*s exhibit shifts in the arm position that feels subjectively straight ahead. This recalibration is of a type that would lead to shifts in subsequent localizing activity compensating for the errors induced by the prism lenses. There is, however, no evidence of any significant shift in the subjective visual straight ahead in this condition. On the other hand, with Terminal Display conditions an opposite pattern of results was found. That is, there was a highly significant compensatory shift in the measure of the visually defined straight ahead and a smaller shift in the measures designed to tap recalibration of proprioception in the exposed arm. Significant adaptation was found in both display conditions as indexed by the measures which involved having *S* manually localize a visible target. This latter type of measure should reflect the cumulative effect of any shifts that manifested themselves in the other two types of measure, and thus this sort of result was anticipated.

The existence of significant proprioceptive-motor shifts in the terminal display condition may seem anomalous, but it should be noted that the model holds attention to be the critical variable in determining the nature of the adaptation developing. The display or exposure condition variable is seen as playing a role only insofar as it influences the source of information to which *S* attends during the exposure period. The relationship between the two is certainly not a perfect one. In view of the evidence regarding what has been termed visual dominance or capture, it might be anticipated that *S*s in the terminal display condition, where exposure circumstances were expected to force attention to proprioceptive cues, would, in fact, attend to visual inputs at least a portion of the time. Though instructed not to make corrective movements when the stylus came into view at the terminus of a pointing movement, no direct control over such activity was possible. To the extent that attention was at times paid to visual cues, adaptation with a proprioceptive locus would be predicted. Some tangential evidence regarding the model's description of the dynamics of the situation can be gained by examining the correlation between the VM and the PM shifts in this condition. Since it would be maintained that adaptation of both types would not develop simultaneously, but rather sequentially as a function of shifts in attention from visual to proprioceptive cues, a negative correlation between the PM and VM shifts would be expected; and, indeed, this correlation is $-.38$ ($p < .05$).

The account given of these findings in terms of the intermodality inconsistency of inputs model has application to an area of controversy in the adaptation literature and can provide an integration and explanation of data which have to this point appeared to be directly contradictory. The issue has to do with transfer of adaptation from a limb observed during exposure to an unobserved limb. Research on this topic has been of interest primarily because of its assumed relevance to some of the same issues dealt with in this paper, specifically, the matter of the locus of the adaptive process. However, Howard and Templeton (1966) have argued that "transfer experiments do not provide an adequate criterion for deciding what is meant by the locus of recalibration along the sensory-motor control loop; one may be able to say only that the relationship between a particular input or set of inputs and a particular output or set of outputs has been changed [p. 381]." They conclude that such investigations enable one to identify the linkages affected but not the loci.

An alternative and, it is felt, more appropriate methodology was employed in the present investigation, and the resultant findings support an interpretation which can account for the discrepant results of this earlier work. That is, the model upon which this study was predicated predicts a visuo-motor locus for the shifts developing when exposure conditions are such that attention is paid primarily to proprioceptive inputs, viz., a terminal display. It seems reasonable that a shift in the direction of gaze is responsible for the observed VM shifts. This sort of mechanism would be expected to manifest itself in shifts in the localization of visually observed targets no matter which limb might be used in making the localizations. Investigations employing terminal display conditions do, in fact, find evidence for interlimb transfer (e.g., Cohen, 1967; Craske, 1967; Howard, 1968). On the other hand, research in which concurrent display exposure conditions were used do not find evidence for such transfer (e.g., Hamilton, 1964; Harris, 1963). The model here predicts a proprioceptive-motor locus for the shifts obtained with a concurrent display and this might well involve a recalibration of the felt position of only the exposed limb, as Harris has argued, such that interlimb transfer would not be evidenced. Cohen found this very pattern, as what he

termed continuous display led to no observed transfer, while terminal visual feedback conditions were associated with interlimb transfer.

The variables of prism orientation and order of presentation of measures included in the analysis of variance were not of direct interest in the study, but rather represented factors it was felt might have an influence on the data and thus should enter into the design as counterbalanced operations. The prism orientation variable was introduced to insure that no extraneous influence such as kinesthetic aftereffects would confound the primary data. Sekuler and Bauer (1966) have shown that repetitive movement of a limb in certain positions can lead to shifts in localization when pointing at visual targets and the significant prism orientation main effect found may well have been due to such a factor. Given the physical characteristics of the experimental situation employed here, the direction of the differences in the magnitude of the shifts found as a function of the direction of the displacement are consistent with the sort of shifts reported by Sekuler and Bauer. The existence of such kinesthetic aftereffects should not, however, have led to any systematic bias in the data of primary interest as prism orientation was a counterbalanced experimental manipulation.

Similarly, the order of presentation of the measures was counterbalanced to insure that neither order dependent nor time dependent effects would confound the data of primary interest. It has, for example, been shown that shifts developing in response to exposure to prismatic displacement can spontaneously dissipate (Hamilton & Bossom, 1964). Little is known, however, about how the time course or the determinants of this decay phenomena relate to the VM and PM measures employed in the present study. Thus it is difficult to speculate regarding the interrelationships of the variables that produced the statistically significant second- and third-order interactions with order of presentation. Research currently in progress was designed to more adequately specify the relationships of time and order-dependent variables as they affect the phenomenon of prismatic adaptation.

REFERENCES

CANON, L. K. Adaptation to a simultaneous displacements of the visual and auditory fields. *Proceedings of the 74th Annual Convention of the American Psychological Association*, 1966, **1**, 13–14. (Summary)

CANON, L. K. Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation. *Journal of Experimental Psychology*, 1970, **84**, 141–147.

CANON, L. K. Directed attention and maladaptive "adaptation" to displacement of the visual field. *Journal of Experimental Psychology*, 1971, **88**, 403–408

COHEN, M. M. Continuous versus terminal visual feedback in prism aftereffects. *Perceptual & Motor Skills*, 1967, **24**, 1295–1302.

CRASKE, B. Adaptation to prisms: Change in internally registered eye-position. *British Journal of Psychology*, 1967, **58**, 329–335.

EDWARDS, A. L. *Experimental design in psychological research.* (3rd ed.) New York: Holt, Rinehart & Winston, 1968.

GIBSON, J. G. Adaptation, aftereffects, and contrast in the perception of curved lines. *Journal of Experimental Psychology*, 1933, **16**, 1–31.

HAMILTON, C. R. Intermanual transfer of adaptation to prisms. *American Journal of Psychology*, 1964, **77**, 457–462.

HAMILTON, C. R., & BOSSOM, J. Decay of prism aftereffects. *Journal of Experimental Psychology*, 1964, **67**, 148–150.

HARRIS, C. S. Adaptation to displaced vision: Visual, motor, or proprioceptive change? *Science*, 1963, **140**, 812–813.

HARRIS, C. S. Perceptual adaptation to inverted, reversed, and displaced vision. *Psychological Review*, 1965, **72**, 419–444.

HAY, J. C., PICK, H. L., & IKEDA, K. Visual capture produced by prism spectacles. *Psychonomic Science*, 1965, **2**, 215–216.

HELD, R. Adaptation to rearrangement and visual spatial aftereffects. *Psychologische Beitrage*, 1962, **6**, 439–450.

HOWARD, I. P. Displacing the optical array. In S. J. Freedman (Ed.), *The neuropsychology of spatially oriented behavior.* Homewood, Ill.: Dorsey Press, 1968.

HOWARD, I. P., & TEMPLETON, W. B. *Human spatial orientation.* New York: Wiley, 1966.

MIKAELIAN, H. Failure of bilateral transfer in modified hand-eye coordination. Paper presented at the meeting of the Eastern Psychological Association, New York, April 1963.

ROCK, I. *The nature of perceptual adaptation.* New York: Basic Books Inc., 1966.

SEKULER, R. W., & BAUER, J. A. Adaptation to prismatic displacements: Hand position and target location. *Journal of Experimental Psychology*, 1966, **72**, 207–212.

(Received April 13, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 240–244

A TEST OF THE FEIGENBAUM AND SIMON MODEL OF SERIAL LEARNING[1]

BRUCE A. FARWELL AND PAUL C. VITZ[2]

New York University

Three groups of 10 *S*s learned serial lists of 8, 12, and 16 nonsense syllables. An analysis of the sequential structure of the learning indicated a number of qualitative weaknesses in the Feigenbaum-Simon (F-S) information-processing model of serial learning. The F-S model predicts *S*s will learn the first two syllables first. But, depending on list length, 40% to 70% of the *S*s learned a later syllable before learning the first two. The F-S model does not allow *S* to learn an item in the middle of the list which is unconnected to the start or end-point chunks. Depending on list length, such isolated syllables were learned by 70% to 100% of the *S*s. The F-S model predicts after the second syllable that learning from the front of the list will be no more efficient than from the back. Instead, *S*s show greater efficiency at forward learning. The significance of the results for the model and the general question of serial learning are briefly discussed.

One of the most stable phenomena in the literature of learning has been the serial position curve, or serial position effect (SPE). The generally bowed shape of the curve, its skewness to the right, and the unevenness of the ends of the curve, indicating more rapid learning of the beginning items, are all well known attributes, yet it has proven extremely difficult to provide an adequate and generally acceptable explanation of them.

Though there have been many hypotheses put forth to account for the various aspects of the SPE, apparently the first attempt to interpret it in quantitative terms was made by Hull, Hovland, Ross, Hall, Perkins, and Fitch, (1940). This model, which required three free parameters, proposed a mechanism of a slowly building ratio of inhibitory to excitory potential with the greatest inhibition in the middle of the list. Several experiments, however, have produced evidence contradictory to Hull's model. (Bowman & Thurlow, 1963; Feigenbaum & Simon, 1962; Glanzer & Dolinsky, 1965; Jensen, 1962).

Atkinson (1957) proposed a stochastic model for the SPE but, as Feigenbaum and Simon (1962) have pointed out, this model has several weaknesses. It assumes four free parameters, thus making it possible to fit the theoretical curve to most data. Furthermore, it has several restricting assumptions: The model holds only for lists of words high in dissimilarity, familiarity, and pronounciability; for moderate presentation rates; and for long (e.g., 1 min.) intertrial intervals.

McCreary and Hunter (1953) demonstrated that all serial position (SP) curves of a given length are essentially identical when the curve is plotted as a proportion of errors at each position. Feigenbaum and Simon (1962) use this to construct an information-processing model which predicts the mean percent of total errors for each serial position in lists of any given length. The Feigenbaum and Simon model is based upon four fundamental postulates, the fourth one being of concern here. This postulate states that the information processes are carried out as follows: *S* considers the ends of the list as "anchor points," learns these first, and then works in toward the middle from both sides. As an important corollary to this postulate, Feigenbaum and Simon specify the sequence of processing and the probabilities involved at each stage: (*a*) The first two items of the list are learned first, and because

[1] This research was supported in part by a grant from the National Institutes of Health, National Institute of Neurological Diseases and Stroke (NS-MH 0736-04A1).

[2] Requests for reprints should be sent to Paul C. Vitz, Department of Psychology, New York University, 707 Broadway, New York, New York 10003.

the model is sequential, these items are learned in the order 1, 2; (*b*) attention is next focused on an item immediately adjacent to one of the anchor points. In the normal list, the number of anchor points is two, these being the third and last items. The probability that an item next to an anchor point will be learned is $1/p$, where p is the number of anchor points. Thus, in the normal list, in which $p = 2$, the probability of learning the third or the last item is $1/2$ for each. (*c*) In the same manner, attention is systematically focused on items adjacent to anchor points until the criterion trial is completed. Essentially, Feigenbaum and Simon assume that after learning Items 1 and 2, the SP curve will be symmetrical from the third to the last item.

Feigenbaum and Simon (1962) then proceed to calculate an SP curve for lists of 12 and 14 syllables according to their model and to compare these predicted curves with the empirical curves of 12- and 14-syllable lists presented in McCreary and Hunter's (1953) study. Using a Kolmogorov-Smirnov test of association for goodness of fit, they found their curves to agree with those of McCreary and Hunter at the 99% level of significance.

Although it is an apparently well-known example of an information-processing approach to verbal learning, the Feigenbaum-Simon (F-S) model has not, to our knowledge, been directly tested. This study specifically investigates three aspects of their model. During pilot work, we observed that *S*s very often failed to learn the second item immediately after the first. That is, the assumption that Items 1 and 2 are learned before going on to any of the later items is suspect. We also observed that *S*s learned items in the middle of the list which were unconnected or isolated from the beginning and end-point chunks. The F-S model does not allow for such isolated item learning, since learning proceeds only from the anchor points. Finally, the F-S model assumes that from the third item to the last, *S*'s learning is symmetric and therefore from Item 3 to the last item the error curve should be symmetric. There are both empirical and rational reasons for doubting this. Ribback and Underwood (1950) found, using a derived list technique, that learning at the front of a list is faster than learning at the back of the list. In addition, the very structure of the serial list implies that learning from the front toward the last item should be more efficient than the reverse. For example, if *S* has learned Items 1 and 2, he is cued at Item 2 that he is at the end of the learned chunk and that he should attend to the next item in order to add to the chunk. However, if *S* has learned items at the end of the list, for example, Items 7 and 8, he must observe Item 7 before he knows that he should have attended to Item 6, which is now past. Thus, in learning from the end of a list to the front, *S* must use his memory of the preceding item in forming the new association.

Method

Subjects.—Fifteen male and fifteen female New York University undergraduate students taking an introductory psychology course were used; the median age was 19.

Procedure.—Nonsense syllables were drawn from Archer's (1960) list of CVC trigrams, with meaningfulness values ranging from 15% to 21%. The CVC syllables selected were assigned to 3 lists, one of 8 syllables, one of 12 syllables, and one of 16 syllables. Ten lists of each length were constructed, using a balanced Latin square design; i.e., no list began with the same syllable and no syllable appeared in the same position more than once. It was necessary, however, to repeat two of the 8-syllable lists in order to get 10 lists. In addition, three 3-syllable practice lists were constructed from CVC trigrams with 35% meaningfulness values. The *S*s were randomly assigned to one of the three conditions. Standard instructions explaining the task were read to all *S*s; after learning the practice list, the experiment began. The *S* responded by spelling each syllable out loud. The list was presented at a constant rate of 3 sec. per syllable. The intertrial-interval was 30 sec. and criterion was one complete trial with no errors.

Results

The obtained serial position curves and the F-S predicted values for the three lists are shown in Fig. 1. The obtained curves are very close to the results of McCreary and Hunter (1953) and Atkinson (1957), although there is a slight tendency

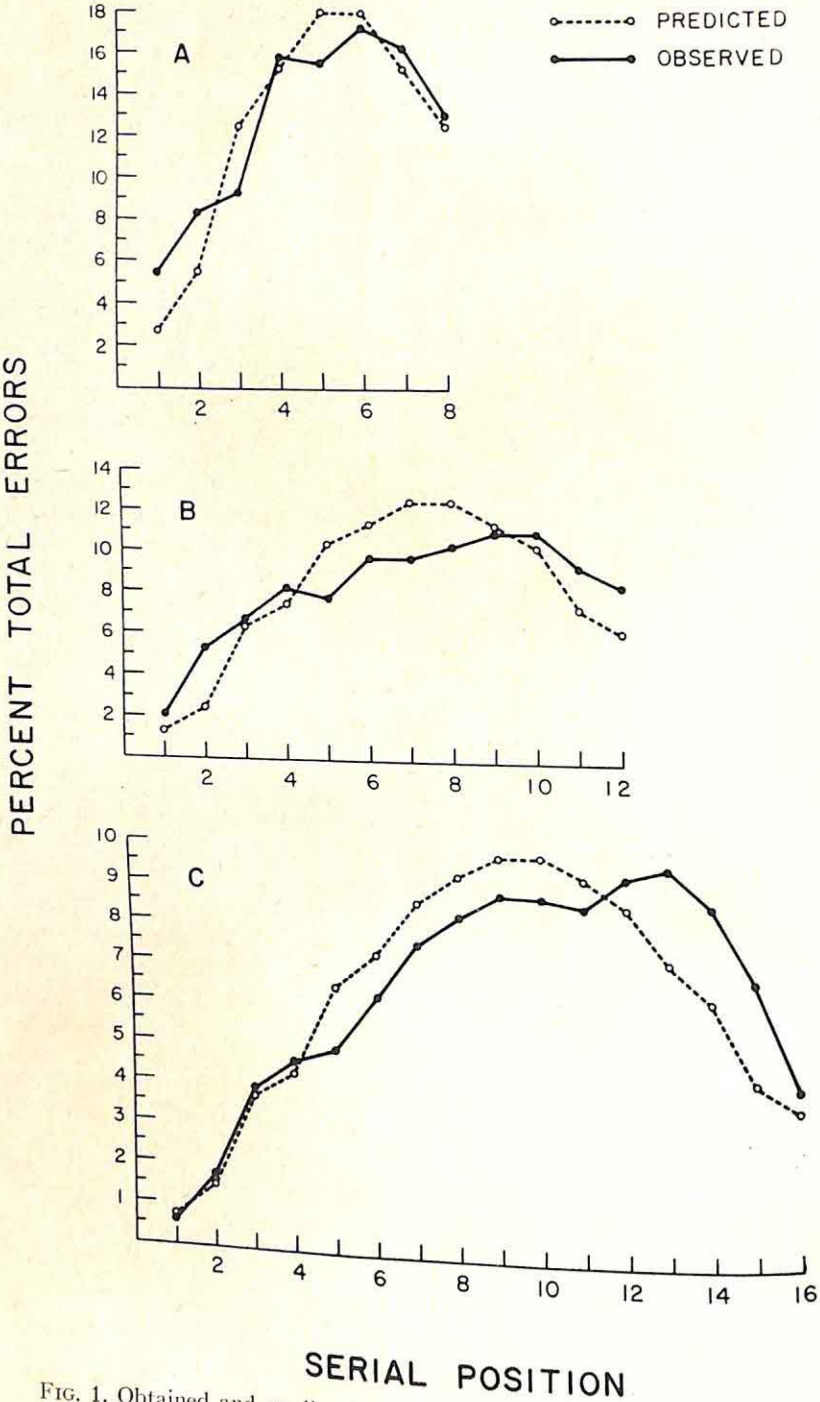

FIG. 1. Obtained and predicted percent total errors as a function of serial position. (Fig. 1A, List 1; Fig. 1B, List 2; Fig. 1C, List 3.)

for our curves to be more skewed to the right. A Kolmogorov-Smirnov test of association (Siegel, 1956) resulted in D values of .054, .048, and .062 for Lists 1, 2, and 3, respectively. These values do not even approach significance and therefore we assume, in all cases, that the fit between the two curves is quite acceptable. A chi-

square test of goodness of fit leads to the same conclusion. By visual inspection, however, it appears that for Lists 2 and 3, the obtained curves are more skewed to the right than predicted; the preceding tests are too weak to respond to such differences. Our main concern, however, is not with exactly how well these curves are predicted. Instead, we are interested in testing the underlying assumptions of the model about how the curves are generated by *S*s' serial information processing.

Table 1 presented the percent of *S*s who learned Syllables 1 and 2 in the order predicted by the model. A syllable was scored as learned if it was anticipated correctly twice in succession without two successive subsequent errors. The trial on which learning was assumed to have occurred was the first trial of the two or more successive correct responses. In Table 1, Column 1 is presented the percent of *S*s who learned Syllable 1 before any other; Column 2 is the percent who learned Syllable 2 second irrespective of what was learned first. Column 3 presents the proportion who learned both syllables in the predicted order. It is clear that for all three lists, *S*s frequently did not meet the assumption of learning Syllables 1 and 2 in that order. Although the SP curve for List 1 fits the model's SP prediction well, the learning of the first two syllables for List 1 is quite discrepant. In the case of the two longer lists, this discrepancy appears to be somewhat less. Because of the small N, the differences related to list length are not significant but there is a tendency for Items 1 and 2 in longer lists to be learned in the predicted order more often than in shorter lists.

TABLE 1

PERCENT OF SUBJECTS LEARNING THE FIRST TWO SYLLABLES IN THE ORDER PREDICTED BY FEIGENBAUM AND SIMON'S MODEL

List and no. of syllables[a]	% of *S*s who learned Syllable 1 first	% of *S*s who learned Syllable 2 second	% of *S*s who learned Syllables 1 & 2 in order
1: 8	60	30	30
2:12	70	30	30
3:16	80	60	60

[a] $N = 10$.

The evidence that *S*s frequently learned an item which was isolated in the list, i.e., items which are not adjacent to previously learned syllables, was also clear. We found in the case of Lists 1 and 2 that 70% and for Lists 3 100% of the *S*s learned at least one isolated syllable.

The main results, shown in Table 2, summarize the evidence that *S*s learn more efficiently from the front of the list (forward learning) than from the back (backward learning). In the case of List 1, the F-S model assumption of symmetric learning from Item 3 to Item 8 is interpreted to mean that on the average, the trial of learning for Syllables 3 and 8 should be the same; likewise, *S*s on the average should learn Syllables 4 and 7, and 5 and 6, in the same number of trials. Similarly, for Lists 2 and 3, Table 2 presents the pairs of item positions which should be learned in the same average number of trials.

In 11 of 15 comparisons, the syllable position nearer the front of the list was

TABLE 2

PERCENT OF SUBJECTS LEARNING ONE OF THE IDENTIFIED PAIR OF SYLLABLE POSITIONS FIRST

List[a]	Syllable position						
1	3 vs. 8 (80) (20)	4 vs. 7 (50) (50)	5 vs. 6 (75) (25)				
2	3 vs. 12 (60) (40)	9 vs. 11 (60) (40)	5 vs. 10 (70) (30)	6 vs. 9 (70) (30)	7 vs. 8 (45) (55)		
3	3 vs. 16 (60) (40)	4 vs. 15 (50) (50)	5 vs. 14 (75) (25)	6 vs. 13 (80) (20)	5 vs. 12 (70) (30)	8 vs. 11 (60) (40)	9 vs. 10 (40) (60)

Note.—Percent of *S*s is given in parentheses.
[a] $N = 10$.

learned first. In two cases, there is no difference and the only two exceptions are small and occur for the two items which are actually adjacent in the list. This split of 11 versus 2 is quite significant, $p = .02$, two-tail, exact probability. When analyzed by *S*s, a similar but less significant pattern emerges. Of the two positions predicted to be of equal average difficulty, those *S*s who more frequently learned the front position earlier than the back position were classed as forward learners. Those who more often learned the back syllable earlier were classed as backward learners. In all, 19 *S*s proved to be forward learners and 9 *S*s backward learners with 2 *S*s that showed no difference, $z = 1.70$, $p < .09$ (two-tail, corrected for continuity).

Discussion

Clearly, the F-S model fails to account for important qualitative aspects of the serial learning process even in the case of the eight-syllable list, where it does fit the SP curve rather well. The main difficulty seems to be the extent to which the F-S learning process is strictly determined. The *S*s are much more variable than the fixed sequence of the model allows. The introduction of a probability distribution across the items could resolve the problem of not always learning Items 1 and 2 first, as well as the problem of learning isolated items, but this would presumably require one or more estimated parameters. The consistent tendency for greater efficiency in forward learning also needs to be represented. If the argument is valid that learning from the end of the list is intrinsically less efficient because of delayed cueing, then this should be part of the structure of the serial learning process.

There is other published evidence which casts doubt on the initial assumptions of the F-S model. Glanzer and Peters (1962) demonstrated that the shape of the SP curve changes with the length of the intertrial interval (ITI). Glanzer and Peters found the SP curve becomes more "peaked" with longer ITI. Therefore, it appears that any satisfactory model of serial learning will need to incorporate in its basic structure at least the following properties: the relative distances among the items, including the ITI, the number of items and a probability distribution across the items, and the greater efficiency of forward learning.

References

Archer, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, **74**, 1–23.

Atkinson, R. C. A stochastic model for rote serial learning. *Psychometrika*, 1957, **22**, 87–95.

Bowman, R. E., & Thurlow, W. R. Determinants of the effect of position in serial learning. *American Journal of Psychology*, 1963, **76**, 436–445.

Feigenbaum, E. A., & Simon, H. A. A theory of the serial position effect. *British Journal of Psychology*, 1962, **53**, 307–320.

Glanzer, M., & Dolinsky, R. The anchor for the serial position curve. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 267–273.

Glanzer, M., & Peters, S. C. Re-examination of the serial position effect. *Journal Experimental Psychology*, 1962, **3**, 258–266.

Hull, C. L., Hovland, C. I., Ross, R. T., Hall, M., Perkins, D. T., & Fitch, F. B. Mathematico-deductive theory of rote learning. New Haven: Yale University Press, 1940.

Jensen, A. R. Temporal and spatial effects of serial position. *American Journal of Psychology*, 1962, **75**, 390–400.

McCreary, J. W., Jr., & Hunter, W. S. Serial position curves in verbal learning. *Science*, 1953, **117**, 131–134.

Ribback, A., & Underwood, B. J. An empirical explanation of the skewness of the bowed serial position curve. *Journal of Experimental Psychology*, 1950, **40**, 329–335.

Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

(Received April 22, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 245–261

VISUAL GUIDANCE OF LOCOMOTION [1]

KEITH R. LLEWELLYN [2]

University of Sydney

The usual conception of visual guidance of locomotion is that it involves the problem of knowing where one is heading. It has been shown that this point corresponds to the center of the radial visual expansion of the surface being approached, and a number of ways in which the expansion pattern might be utilized in detecting this point have been suggested. This conception of the problem is rejected and it is argued that guidance situations requiring accuracy involve locomotion toward a specific target. The perceptual information for this guidance is provided by the movement of this target, the presence and nature of any drift signaling the presence of a heading error and the nature of the correction required. The rejection of expansion information judgments as the probable basis of guidance is based on the lack of response variation with variation of the stimulus display (at times extreme) and the high degree of inaccuracy consistently displayed. No improvement in accuracy occurred with variation of the viewing conditions, nature of the displays, and nature of the task, or from the use of monocular or binocular vision, free viewing or fixation, sophisticated *S*s, and extensive instruction in the nature of the expansion pattern. By comparison, the detection of target drift was shown to be affected by stimulus variation, the presence of an expanding background reducing efficiency of detection. Accuracy of cancellation of target drift was shown to vary with stimulus variation but to remain at a level of accuracy that was much superior to that of judgments based on expansion information.

The problem of how locomotion is visually guided is commonly expressed in the form, "How do we know where we are going?" The answer to this question has been seen as requiring the identification of the physical characteristics that uniquely specify the particular point in the environment being approached. The assumption is then that, as it is obvious that visual guidance can reach a remarkably high level of accuracy, *O* must be able to react to the physical characteristics of the point being approached and detect its location with a very small degree of error. The independent analyses of Gibson (1950) and Calvert (1950) have indicated the nature of the physical uniqueness of this point. They have shown that as a surface is approached or traversed, it expands (in the sense that the visual angles subtended at the eye by features of the surface increase in size). This expansion takes the form of a radial movement of these features away from the point being approached, which is therefore characterized by being the center or focus of the optical expansion. For convenience, Calvert's usage will be adopted and the center of expansion referred to as "X."

If the information necessary for visual guidance is provided by the expansion pattern, it should be possible to improve judgments in critical guidance situations by providing textures which would clearly define the expansion pattern and its focus. In order to choose the pattern that does this most efficiently, a parametric study of the effect of textural variables should be carried out. The first two experiments reported here are of this parametric kind. The results were entirely unexpected, partly in that the errors made by *S*s were extremely large. If these results were to be accepted, the implication would be that *O* cannot tell with any accuracy where he is heading. This conclusion seems so obviously absurd in the light of the accuracy

[1] Experiments I to VIII formed part of a PhD thesis submitted to the University of Sydney, 1962. Appreciation is extended to the Department of Civil Aviation and to the Department of Supply for the provision of a grant and equipment. Experiments IX and X were subsequently carried out at the University of New South Wales.

[2] Requests for reprints should be sent to Keith R. Llewellyn, who is now at the University of New South Wales, School of Applied Psychology, Kensington, Australia.

of visual guidance that a thorough examination would have to be carried out to reveal why such unsatisfactory results were obtained. The results of such an examination will be presented in Exp. III, IV, V, and X. To anticipate these results, the conclusion is drawn that it is highly probable that the large errors are not specific to the experimental situation but are typical of the judgment, so that in fact *O* cannot tell where he is heading with any accuracy.

Such a conclusion demands a reexamination of the whole question of visual guidance. It would seem that *O* does not need to know where he is heading with any degree of accuracy if he is freely locomoting and not restricted to guiding himself toward a particular goal or target such as the next portion of roadway or the threshold of a runway. In such cases, a general sense of direction provided by the visual expansion, supplemented as it usually is by orientation cues from the body or vehicle, is probably sufficient. If the locomotion is directed toward a particular target, the accuracy or inaccuracy of *O*'s knowledge of where he is heading may be almost irrelevant. Visual guidance may depend upon some quite different judgmental task. The use of expansion pattern information in the detection and correction of heading errors would consist of the detection of the location of the center of expansion (X), recognition of its noncorrespondence with the target, and adjustment of the path of locomotion so that X is moved until it coincides with the target. An alternative hypothesis makes the detection of X unnecessary: *O* merely has to know (*a*) that he is not moving directly toward the target and (*b*) that he may correct this situation by making certain alterations to his path. The physics of the situation must remain the same—X is moved to coincide with the target—but the perceptual information controlling this adjustment would derive not from the expansion pattern or X but from the movement of the target itself. Considering the target as being located on a picture plane in the fronto-parallel, so that movement of the target toward *O* is represented by an increase in target size, the presence of any target movement other than the spreading of the contours signals a heading error. If the target drifts to the left, for example, *O* is off-course to the right and must correct his path of movement in the direction of the target drift so that the drift is canceled. If the rate of drift is relatively high, the heading error is large and a large course correction is required.

Visual guidance of locomotion is therefore not synonymous with knowing where one is heading, except in the special case of knowing that one is heading toward a desired target. When this is not the case, *O* might have only the vaguest of notions as to where he is actually heading—"over to the right somewhere"—but still be able to correct for the heading error by canceling the drift of the target.

Gibson (1966) has moved somewhat in the direction of this viewpoint in his later writings in which he distinguishes between approach to a surface and approach to an object. Thus he says, "Visually guided locomotion is a matter of going to a specific goal in the environment [Gibson, 1966, p. 162]" and, in discussing animal locomotion, says that all the animal has to do is to magnify the form of the desired goal in order to reach it. Such a procedure is a very crude basis for visual guidance and Gibson still contends that "The same rule of visual approach holds true for swimming, flying or running: keep the focus of centrifugal flow centered on the attractive thing or the inviting place [1966, p. 162]."

The task of detecting and canceling target drift is much less demanding of the stimulus situation than the tasks specified in the various expansion-pattern theories. Gibson's theory seems to require the presence of an extended differentiated surface which would define the pattern of relationships between velocities (in his 1950 account) or between shape transformations (in his 1966 account). Calvert's theory is less demanding in this respect as he considers that the "streamer pattern" is provided by the paths of movement of distinctive objects such as the corners or stub-bars of the runway or the trees and

houses that might be near the airfield. These objects might be few in number and indefinite in shape, for example, runway lights at night. Calvert is more restrictive than Gibson in his viewing requirements as he specifies that *O* should fixate a target so that the movement information from the streamer paths might be detected by the parafovea. Another possible use of the expansion pattern is the weighting of the velocities of objects moving in different directions, and one that does not require the expansion pattern as such but implies the existence of numerous surface elements is the detection of an element that is not moving and therefore corresponds to X. All these theories require the presence of some form of differentiated surface, and consequently they cannot account for guidance toward a target if no background is present, for example, flying toward an enemy aircraft seen against the clear dome of the sky or toward a single lighted target at night. By comparison, the alternative theory is that locomotion toward a target requires the presence of a target, and nothing more. Indeed, the presence of a visually dominant expanding background to the target could detract from the accuracy of the guidance by making drift harder to detect.

The initial parametric studies did not assume any particular method of utilizing the expansion pattern. Instead, it was assumed that *O* would automatically adopt the method appropriate to this highly practiced task. The investigation of possible reasons for the very large errors that occurred in these studies included experiments specifically designed to provide appropriate conditions for judgments based on Gibson's hypothesized use of the expansion pattern (Exp. IV and X) and for Calvert's hypothesized use (Exp. V and X).

The critical examination of the results of the first two experiments led to the conclusion that they were unreasonable because they failed to show two properties which are necessary if a particular judgmental task is to be considered the probable basis of visual guidance. These two requirements were therefore expressed as general criteria for the assessment of the results of all subsequent experiments. These criteria are that (*a*) as visual guidance is normally carried out with accuracy under a wide range of stimulus conditions in extralaboratory situations, the responses in the particular laboratory situation used should not be markedly inaccurate; and (*b*) despite this general tolerance of the nature of the stimulus situation, the responses should clearly be a function of the visual stimulus information presented. In other words, some variation in the generally high level of accuracy should occur as, for example, texture density is varied over a wide range or as the texture is changed from a pattern of dots to a grid composed of straight lines.

These criteria are basic, so that the demonstration that one type of judgment is more accurate than another is no evidence that the more efficient judgment is the probable basis of visual guidance if its error rate is high and it shows no sign of being dependent upon what is hypothesized to be the relevant stimulus features. The application of these criteria therefore constitutes the main concern of these studies.

General Method

The *S* sat in an enclosed observation chamber and viewed a screen on which a shadow pattern was displayed. The pattern was made to expand radially from a center of expansion in a manner simulating the expansion of a surface approached at an angle of 90°. The textural characteristics of the shadow pattern were varied and the position of X could be varied from trial to trial.

Apparatus.—The apparatus was a shadow projector as described by Gibson (1957). A patterned sheet of glass was drawn toward a point source of light and the shadow of its pattern formed an expanding image on a translucent screen viewed by *S* from the opposite side. The position of the light could be adjusted in the plane parallel to the screen, and as the intersection of the screen and the perpendicular to it from the light corresponded to X, such an adjustment allowed *E* to position X anywhere on the screen. Relevant details are given in Fig. 1 and Table 1. The responses showed a high degree of consistency despite changes in the particular light source used and even despite the use of apparatus in the last two experiments which was of the same principle as that used earlier but quite different in detail (for example, the screen was composed of different material). It is therefore clear

TABLE 1

Details of Experiments

Variable	Specification	Exp.
Light to screen distance	231 cm.	I–VIII
	240 cm.	IX, X
Rate of movement of patterned glass toward light	3.6 cm/sec	I–III, V–VIII
	3.7 cm/sec	IX, X
	5.4 cm/sec	IV
Initial light to glass distance	131 cm.	VI
	156 cm.	IV, V–VIII
	161 cm.	I–III
	194 cm.	IX, X
Final light to glass distance	47 cm.	I–III
	51 cm.	IV–VIII
	61 cm.	IX, X
Size of visible screen	Round, 56 cm. diameter	I–III
	76 × 76 cm.	IV–VIII
	71.5 cm. high × 65.5 cm. wide	IX–X
Viewing cond.	Field stop, round aperture about 10 cm. from eyes, binocular	I, II
	Field stop, round aperture about 10 cm. from eyes, monocular	III
	Headrest, unrestricted binocular vision	IV–X
S to screen distance	100 cm.	All

that the basic requirements are that the display should be reasonably well-defined and should cause no visual discomfort to *S*.

Texture patterns.—Dot patterns were generally prepared by allowing small drops of paint to flow out of a stencil pen on to the surface of a sheet of plate glass. The resulting dots, if not rejected and replaced, were roughly circular, approximately one-sixteenth of an inch in diameter and evenly and densely opaque to their edges. Larger dots were prepared by punching circular discs from thick glue-backed paper or from plastic insulation tape and these were glued in position. Patterns containing lines were drawn with Chinagraph pencil, except for the thick-lined grid used in Exp. II in which thick strips of paper were glued on the glass. Extended patterns covered an area of at least 63 × 63 cm. so that no edge to the pattern was visible on the screen.

Some of the small dot patterns were designed to give the impression of random distribution while the actual dispersion of the dots was fairly strictly controlled. For example, one pattern, with a density of 25 dots per square inch, was constructed by preparing a template 4.4 × 4.4 in. One dot was allocated to each .2 × .2 in. square on the template, the irregularity being achieved by dividing these small squares into four subunits each .1 × .1 in. and randomly allocating the dot to one of these subunits. The template was then used to produce the large field of dots, detection of the repetition of its pattern being successfully prevented by randomizing its orientation (Two Possible Faces × Four Possible Sides) for each application. Such patterns will be referred to as "irregular."

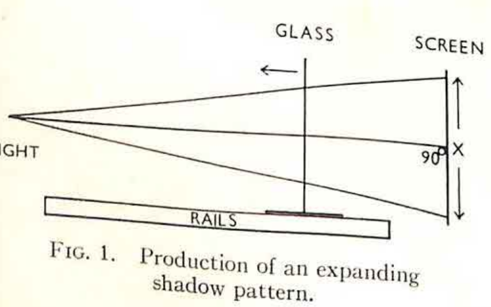

Fig. 1. Production of an expanding shadow pattern.

Evidence against Expansion-Information Theories: Initial Parametric Studies

Experiments I and II

The first experiment examined the textural variable of regularity, two regular patterns (small dots in rows and columns) being compared with two irregular dot patterns. Density was also varied by having one of each type of pattern with a density of 25 dots per square inch (about 16,000 dots visible at the start of the trial) while the two remaining patterns had a density of 6.25 dots per square inch (about 4,000 dots visible). The same plan of element placement was used for the two irregular patterns so that in one sense the 6.25-dots-per-square-inch pattern was an expanded version of the denser pattern.

In order to prevent *S* from remembering judgments and adopting stereotyped responses, the position of X was varied. Two positions were used in the demonstration trials and four other positions in the test trials (Fig. 2).

Method

Subjects.—Twenty-four students from the psychology course were used. They varied from first year to postgraduate, but all were unfamiliar with the research topic.

Procedure and design.—The *S* was given a set of instructions which contained the following description of the nature of the optical expansion:

> The dots will grow larger and will move across the screen away from the center of expansion, that is, if the dots left a trail behind them, these trails would all radiate from one section of the screen. The centre of expansion can be located anywhere that you can see. . . . If a dot happens to coincide with the centre of expansion it will not move across the screen but will simply grow larger.

The S was given 2 demonstration trials followed by 16 trials in which each pattern was presented for a block of 4 trials. Each of the 24 possible sequences of the four pattern blocks was randomly allocated to one of the 24 Ss. The 4 trials within a block consisted of four different positions of the center of expansion. The same four positions were used for each block, but the sequence of positions was randomized.

The S indicated the judged location of the center of expansion at the end of each trial by positioning a spot of light on the screen. This was produced by a torch which was on only during the response period. The orientation of the torch was read off a dial outside the observation chamber.

Results

Experiment I.—The responses were recorded in terms of angular coordinates (torch orientation) and these were converted to plots on a chart of the screen. The discrepancies between the judged and correct positions of X were measured as linear deviations, rounding the result to the next highest quarter of an inch. An analysis of variance performed on these error scores yielded only one significant effect, that of position of the center of expansion, $F(3, 69) = 5.091$, $.05 > p > .01$, indicating that central positions of X were the most accurately judged.

Experiment II.—The second experiment followed the same basic procedure and examined the effects of variation of the nature of the elements of the pattern and of their size. The patterns used were a regular small dot pattern, with a density of 6.25 dots per square inch, a pattern having the same regular dot arrangement and a very similar density, but having dots of $\frac{1}{4}$ in. in diameter instead of $\frac{1}{16}$ in. and two patterns in which solid lines replaced the rows and columns of dots. The thickness of the lines was matched to the diameters of the dots. The X was varied over seven positions (Fig. 3) and the seven Ss had all taken part in the first experiment.

The main interest was in the grid versus dot element variation, for patterns of radial movement might be expected to be much easier to detect with discrete, isolated units than with an integrated pattern. On the other hand, the pattern of shape transformations presented by the enclosed squares of the expanding grids might give those patterns superiority.

The results were similar to those of the first experiment, the only significant result being the main effect of the position of X, $F(6, 36) = 5.89$, $p < .001$. Once again, the central position was the most accurate.

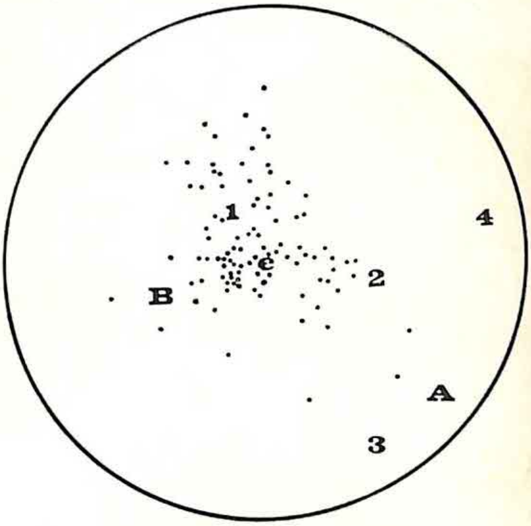

FIG. 2. Location of two practice positions of X (A and B) used in Exp. I and the four test positions; responses to Position 1, all patterns. (C = center of screen.)

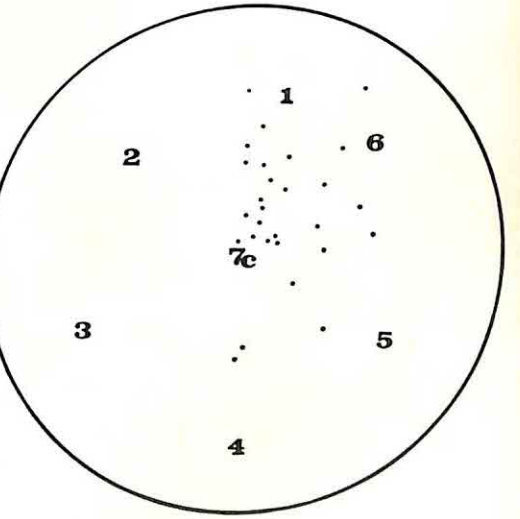

FIG. 3. Location of the seven test positions of X used in Exp. II and responses to Position 1 for all backgrounds. (C = center of screen.)

Discussion

The responses in these two experiments failed to satisfy the criterion that some variation should occur with marked changes of the stimulus presentation. This failure was not due to the tolerance of a highly practiced skill of such stimulus changes, as the responses definitely fail the error-size criterion. The mean errors for the 16 pattern/position conditions for Exp. I ranged from 3.01 in. to 5.24 in., corresponding to visual angles from 4°18′ up to 7°25′. The raw scores ranged up to 15 in. (on a screen only 22 in. in diameter). The same large mean errors were evident in the second experiment, and the raw scores ranged up to 11.5 in.

The significant result due to position of the center of expansion was affected by two artifacts in the situation. With errors of the size obtained, the variation in the maximum possible error becomes important. The maximum error for a central position would be about 11 in., the radius of the screen, while that for a peripheral position would approach 22 in. Another artifact was the marked tendency for the "center of gravity" of the responses to be toward the center of the screen (Fig. 3). Despite this, the responses for central positions were widely scattered, the responses for the most accurate position in the first experiment covering half of the screen (Fig. 2).

Evidence against Expansion-Information Theories: Examination of Possible Artifactual Causes of the Large Errors

Experiment III: Random Error from an Unknown Source

If the systematic effects of stimulus variation were being swamped by random error induced by some unknown aspect of the situation, they should be revealed if the stimulus variation is extreme. This was tested by using two patterns that differed in regularity and density. The regular pattern presented an initial display of about 4,000 dots in rows and columns and a final display of about 95 dots. The irregular pattern had 130 dots visible at the start of the trial and only 8 at the end.

Viewing was changed from binocular to monocular to avoid focusing difficulties that may have increased the difficulty of the task. The 20 *S*s were given 4 demonstration trials, using two positions of X (A and B) and 36 test trials, the 18 trials for each pattern consisting of 3 for each of the six positions of X. Two positions of X were to the right of the screen, two to the left and two near the center. All were near eye level.

The results were examined in detail and typical results are shown in Table 2. As the results for the two patterns were almost identical, it seems unlikely that systematic effects were being swamped by random error.

Experiment IV: Inadequate or Inappropriate Instructions

An implicit assumption has been that although few persons would be aware of the expansion pattern while locomoting, nevertheless *S*s could consciously attend to it and make use of it when instructed to do so in the laboratory. This is probably the case with possible sources of information for other aspects of locomotion, for example, the rate of flow of roadway joints toward a vehicle. If the expansion pattern were of this class of perceptual information, the inaccurate responses could have resulted from inadequately instructing *S* as to the rather complex aspects to which he had to attend. To test this possibility, an expanded set of instructions was prepared. This took the form of an individual lecturette in which the nature of the expansion pattern was fully explained and illustrated by means of diagrams. Instructions of this kind would be appropriate to most of the possible ways in which *S* might make use of the expansion pattern to locate X.

By this stage in the research, the reports of *S*s and the observations of *E* had indicated that theories assuming that *S* could directly perceive the radial nature of the expansion pattern were most unlikely to be correct. One can attend to a limited portion of the pattern and perceive direction of movement and, to some extent, rate of movement, but when the whole pattern is attended to the differences in the velocity

and direction of movement of elements are almost nondetectable. There is a strong illusory impression of the movement of the pattern toward *O*, and this illusion is resistant to knowledge of the fixed nature of the screen. The impression is of a unitary surface coming forward in a single direction rather than of a collection of elements which move in many directions. The expansion of the pattern appears uniform and seems to be limited to an awareness that the distances between elements are increasing, and it would seem that even this impression may disappear when the boundaries of the visual field are made sufficiently indefinite, giving the effect of a pattern of fixed size coming forward.

Gibson's theory is concerned with the overall pattern of change of the surface and he has indicated that *S* could be instructed to use the expansion pattern to locate the "focus of expansion" (1950, p. 130). Despite this, it would seem that the informal data should not be used to dismiss Gibson's theory as it seems likely that it does not depend upon the direct perception of the expansion's radial flow (statements by Gibson subsequent to this experiment support this view). The expansion pattern should be regarded as a "stimulus correlate," which would mean that instructions emphasizing the physical expansion pattern would be inappropriate as they would be requiring *S* to perform a task analogous to judging the degree of his retinal disparity rather than one analogous to judging the distance of an object. The instructions would have to emphasize the end product, the awareness of the point toward which one was heading, and instructions of this type were therefore administered to a second group of *S*s, the "Approach" group.

The primary aim of this experiment was not to compare the performance of groups having different instructions but to try to find an instruction group that gave responses that could be described as being "reasonably accurate."

TABLE 2

RESULTS FOR BOTH CENTRAL AND ONE OF EACH LATERAL POSITIONS

Position and pattern cond.	Results (in in.)			
	Mean error	Range of errors	Displacement of mean position	Distance between mean positions
Left regular	6.4	.5 –17.0	5.5	.3
Left irregular	6.8	.5 –20.0	5.8	
Center regular	3.6	.5 –11.0	3.5	
Center irregular	3.4	.5 –10.0	3.2	1.3
Center regular	3.9	.5 –10.0	1.7	
Center irregular	3.3	.5 – 9.0	2.2	1.3
Right regular	3.4	.5 –13.0	2.4	
Right irregular	3.6	.75–13.0	2.3	.2

Method

Apparatus.—Two changes were made to the apparatus to make the task easier. The field stop was removed and replaced by a headrest. This allowed *S* to use binocular vision to obtain an unrestricted view (about 30 × 30 in.) of the whole screen and of the surrounding framework. The speed at which the patterned sheet was drawn toward the light was increased from 3.6 to 5.4 cm/sec.

Procedure and design.—For each group of 14 *S*s, each *S* was given the appropriate instructions and one demonstration trial, followed by 10 test trials consisting of 5 trials on each of two positions of X. Position 1 was at the center of the screen and Position 2 was 20 cm. to the left at the same level. The demonstration position was midway between the two. The order of positions over trials was randomized and the same stimulus display, the irregular small dot pattern density 6.25 dots per square inch, was used for all trials.

At the end of the test session, *S* was questioned about his comprehension of the instructions and the difficulty of the task.

Instructions.—The relevant section of the Approach instructions was:

> The pattern will start to look as though you are moving towards it. The edges of the screen will look as though they are the edges of a window. You are looking through the window at the dots. What I want you to do is . . . to try to locate the exact point towards which you are moving.

The Radial *S*s were each given a lecturette about the nature of the expansion. A picture of the initial appearance of a dot pattern was shown, followed by one of the expanded final appearance. A third showed the initial appearance superimposed on the final, with some of the radiating paths of movement indicated and the location of X marked. A final picture showed a similar composite with X towards the edge of the pattern instead of near the middle.

TABLE 3
ERRORS OF THE TWO INSTRUCTION GROUPS

Group and position	Errors (in in.)		
	Mean error	Range of errors	Dispersion error
Approach 1	4.43	1.00–12.50	3.80
Approach 2	6.16	.50–18.50	5.15
Radial 1	6.54	.25–18.50	5.21
Radial 2	6.36	1.00–19.00	5.54

Note.—Dispersion error is the mean absolute deviation from the mean judged position for each trial, averaged over trials.

Results

Responses to questions.—All *S*s indicated that they understood the instructions. One member of the Approach group said that they could have been more detailed. In general, the Approach group found the task undemanding and had reasonable confidence in their judgments, while the Radial group found it demanding and had little confidence.

Judgments of the correct positions: (a) Group results.—The results presented in Table 3 indicate that neither group could perform the task with reasonable accuracy. The only mean error less than 6 in. was for the central position by the Approach group whose responses tended to be more definitely drawn towards the center of the screen. The dispersion errors (deviations from the mean judged positions) show that the responses were not clustered around a displaced mean judged position.

Judgments of the correct positions: (b) Individual results.—The possibility that the accurate results of some *S*s were being obscured by the highly inaccurate results of others was examined by setting criteria for "accurate and meaningful" responses. These were that *S*'s mean error for a position over the five trials should be 3 in. or less, that the mean deviation over the first two trials should not be exceeded by the mean deviation from X over the last two, and that these two criteria should hold for both positions. None of the Radial group qualified for the first criterion and only three of the Approach group qualified for Position 1 and three for Position 2. When the second criterion is added, 2 of the 3 *S*s failed to qualify for Position 2, and only 1 *S* out of 28 satisfied all three criteria.

Discussion

Detailed instruction in the nature of the expansion failed to lead to any improvement in performance. Changing the nature of the task by emphasizing judgments of the point being approached likewise produced no improvement in the responses, so that it would seem that the large errors obtained in the previous experiments were not due to inadequate or inappropriate instructions.

Experiment V: Inappropriate Displays and Lack of Fixation

The displays used may have been inappropriate for examining theories which emphasize the use of a limited number of surface features as they were generally composed of many identical elements distributed in a fairly uniform manner over the whole screen (but see the irregular pattern in Exp. III). Calvert's theory (1954), for example, differs from that of Gibson (1950) by emphasizing these noticeable features, probably as a result of Calvert's concern with aircraft landings at night and of his awareness that the landing strip and its surrounds often have minimal texture. The surface as such would form a generally undifferentiated background to the distinctive features of the strip (its corners and stub-bars) and to the houses, trees, and natural formations in the vicinity. Calvert's theory also differs from Gibson's in its identification of the direction of movement of these features as the basis of guidance rather than patterns of velocities or shape transformations. For Calvert, the location of X is given by the intersection of what would be the backward extension of the paths of movement of these features (the "streamers") if the locomotion were in a straight line towards the surface. In order to take account of the complex paths of movement resulting from the maneuvering of the aircraft, Calvert specifies that X is the "instantaneous intersection" of the tangents to the streamers. The direction of movement of these streamers is said to be detected by the parafovea, and Calvert

emphasized that *O* fixates so that this perceptual analysis can be carried out.

It was decided to test Calvert's theory by using a limited pattern (an idealized representation of a runway) and providing for fixation. As Calvert stated that the pilot fixates his target, a small target was incorporated in the pattern near the threshold of the "runway." The runway pattern was designed to provide parafoveal stimulation and it would provide the multiple streamers necessary for the specification of the location of X, the intersection of the streamer tangents.

This visual presentation may be contrasted with one in which the crucial aspects—parafoveal stimulation and multiple streamers—are absent. Such a display would be provided by the fixation target if presented alone, provided that it was circular with no noticeable irregularities and was small enough to remain limited to the fovea. As the visual angles quoted for foveal and parafoveal stimulation vary from author to author, it was decided that the runway would be deemed to provide parafoveal stimulation if it extended beyond 5°, the largest figure given for foveal stimulation, while the target would be deemed to provide only foveal stimulation if it never exceeded 1.5°, the smallest angle given.

Method

Visual displays.—The parafoveal display consisted of an upright rectangle with short horizontal projections midway up the long sides. At the start of the trial, the height of the rectangle corresponded to 5°20′ and its width to 2°14′. The projections increased the width to 2°48′. At the end of the trial, the height corresponded to 16°18′ and the width to 6°52′, or 7°26′ with projections. The target, placed slightly above the center of the base of the rectangle, had a diameter corresponding to 0°8′ at the start of the trial and 0°24′ at the end. The X was 14.4 cm. from the target at the start of the trial and the target moved 29.6 cm. down and to the left, at an angle of 60° to the horizontal from a position at eye level and midway across the screen.

The foveal display consisted of the same target with the rectangle and its projections removed.

Subjects and procedure.—Two groups of 20 *S*s (students and clerical and junior academic staff) were used. They were told to try to imagine that they were in the cockpit of an aircraft which was approaching the screen display and that they were to judge the point being directly approached while fixating the target. They were then given two demonstration trials followed by five test trials. All Parafoveal group *S*s were tested with the rectangle present, and this was then removed and all Foveal group *S*s tested.

TABLE 4

ABSOLUTE DEVIATIONS FROM X

Group	Errors (in in.)			
	Mean error	Grand median	Median splits	
			−	+
On fifth trial				
Foveal	6.38	5.95	9	11
Parafoveal	5.87		11	9
Over five trials				
Foveal	6.17	5.37	9	11
Parafoveal	5.92		11	9

Results

The most noticeable feature of the results was the extreme inaccuracy displayed by both groups. The Foveal group errors ranged up to 15.5 in. and those for the Parafoveal group up to 14 in., and the mean errors were very large. The errors were compared by splitting their distributions at the overall median and, as Table 4 shows, the Parafoveal group was obviously no less inaccurate than the Foveal.

TABLE 5

LOADINGS ON A AND B AXES

Group	Loadings (in in.)	
	A	B
On fifth trial		
Foveal	1.11	−.82
Parafoveal	−.14	−5.06*
Over five trials		
Foveal	.27	−2.00
Parafoveal	.05	−4.59*

* $p < .001$.

A plot of the responses on a chart of the screen showed an apparent difference in their distribution. This was examined by drawing a set of axes, A and B, with Axis B lying along the path of movement of the target and the intersection of the axes at X. Two criterion scores were obtained for each axis—loading on the fifth trial and mean loading over the five trials—and Table 5 shows that the parafoveal responses were significantly biased away from X toward the final position of the target for both measures. The Foveal group showed no such systematic bias in their dispersion, and the difference between the groups was significant when the criterion was the response on the fifth trial ($.02 > p > 0.1$).

Conclusions

As the presence of those features of the expansion pattern said by Calvert (1954) to be the basis of detection of X, namely, multiple streamers in the parafoveal region, produced responses that were just as inaccurate as those made in the absence of these features, and as the only difference between the conditions consisted of a systematic bias of the parafoveal responses away from the position of X, it was concluded that Calvert's theory should be rejected.

Other Possible Causes

Two possible causes of the large errors would be the restricted size of the screen and the use of binocular vision. Before evaluating these features, it is necessary to distinguish between the particular aspect being investigated—guidance—and another aspect of locomotion, the subjective impression of one's motion toward the surface. As no theory of guidance specifies that this subjective impression of motion is necessary for accurate guidance, the requirements of this effect (for example, Gibson's (1954) requirement that the expansion must cover the whole visual field) would not seem to be relevant to the present research.

Gibson has actually championed the use of binocular viewing of shadow caster displays to study movement effects (Gibson & Gibson, 1957); and as the use of monocular viewing in Exp. III produced no improvement in the responses, it is doubtful if this factor could be responsible for the large errors. The screen was of restricted size, but in Exp. IV, for example, it would have corresponded to a windscreen about 15 × 15 in. in size at a distance of 20 in. As this "windscreen" was filled with the expanding pattern and did not have an almost textureless "sky" occupying a large section, it would not be an unreasonable representation of the visual field likely to be present in many forms of vehicular locomotion. In any case, if the size of the screen was inadequate for the use of the expansion pattern, a judgment requiring extended fields of view is not likely to be of much use when one is driving at night or flying a fighter with restricted forward vision.

In normal locomotion, surfaces generally maintain a fairly constant density despite the expansion as new elements become visible and join the flow pattern. None of the patterns used had this feature. If this were relevant, one might expect differences in response accuracy to occur when texture density variation is extreme (Exp. III) and reasonably accurate responses to be given when the density is high (Exp. I). Displays of lights approached at night also do not have this property and it does not seem likely that the guidance of such an approach would involve a radically different perceptual task to that towards an extended surface in daylight.

The rate of simulated approach was intended to provide a definite but not explosive rate of expansion on the screen. Palmer (1969) has systematically varied the rate of simulated approach and found that although errors are smaller when faster expansions are used they still remain unacceptably large. He also found that variation in the simulated angle of approach (90° and 15°) had no effect on the responses.

It is considered that even if some deficiency in the laboratory situation (such as the limited size of the screen) is responsible for the very large errors, any judgment that requires optimal conditions for accuracy is unlikely to be the basis of visual guidance. Such conditions are not always present in common guidance situations and it seems uneconomical to conclude that the basis of guidance differs in such situations from that in less restricted conditions.

Evidence Supporting the Target-Drift Theory: Factors Affecting Detection of Target Drift

Experiment VI: Drift Detection with and without an Expanding Background

It has been suggested that instead of carrying the relevant perceptual information, the expanding background could serve to reduce the efficiency of guidance by making target drift harder to detect. The first experiments on target drift were mainly concerned with this role of the background and provide evidence that the task satisfies the criterion of varying in efficiency as the stimulus conditions are varied.

Eight *S*s were tested with a target presented either by itself or against a grid background (the thin-lined grid of Exp. II). Both target and grid were made to expand, the target increasing from a diameter of .9 in. to one of 2.3 in. The *S*s were required to detect the presence of the drift produced by displacing the target to the left or right of X. The displacements were 0, 1.18, or 2.08 in. and the target was 0, 3.03, or 5.35 in. from X at the end of the trial. The major response measure was latency of detection of drift, and the use of a different measure on half of the trials gave evidence that these measures were valid. An analysis of variance of the latencies showed that the mean latency with the grid present (16.73 sec.) was significantly greater than the mean latency when it was absent (13.94 sec.). The use of different displacements provided evidence for the basis of drift detection. The significantly different mean latencies (13.05 sec. for the large displacement and 17.63 sec. for the small) can be converted into distance moved by the target. As the velocity of the target was determined by its distance from X, the detection of movement should take place at much the same X to target distance regardless of the initial size of this distance if detection depends upon the target reaching a threshold velocity. These distances were 6.49 cm. and 9.25 cm., respectively, for the small and large initial displacements with the grid background and 5.82 cm. and 7.96 cm. for no background. The actual distances moved by the target, 3.49 cm. and 3.96 cm. for the grid present and 2.82 cm. and 2.67 cm. for the grid absent, were much more constant, suggesting that distance moved rather than velocity attained underlay detection.

Experiments VII and VIII: Drift Detection with Quantitative Stimulus Variation

Method

Having shown that the detection of drift varied in difficulty with a qualitative change in background conditions, an attempt was made to discover the effects of quantitative changes in both the background and the target. The background chosen was the irregular dot pattern used in the early experiments and the density of this was varied over a number of steps. The size of the target was also varied. Details are given in Tables 6 and 7.

For Exp. VII, 24 *S*s were presented with four background conditions in blocks of nine trials. The nine combinations of three target sizes and three movement conditions (left, right, and no movement) were randomly ordered within each set of nine trials. Each of the 24 possible orders of background was allocated randomly to an *S*. The response measure was latency of detection of drift.

Results

An analysis of variance of the latencies yielded significant F ratios for density and target size (F (3, 69) = 11.96, $p < .001$, and F (2, 46) = 3.82, $.05 > p > .025$. An examination of the means for density (Table 6) showed that the mean latencies for very different backgrounds (3 and 4) were remarkably similar and that the change to Background 2 was only slight. The major effect appeared to be produced by the introduction of a background as opposed to none, and post hoc comparisons confirmed that the only significant differences were between the conditions of no expanding background (1) and all other background conditions.

Background 2 had not been prepared as carefully as the other backgrounds and its average density was found to vary from

TABLE 6

BACKGROUND DENSITIES AND MEAN LATENCIES

Display	Background densities (in dots per square inch)					
	1	2a	2	2b	3	4
On glass slide	0	.25	.57	3.25	6.25	25.00
On screen at start of trial	0	.11	.25	1.44	2.78	11.11
On screen at end of trial	0	.01	.03	.16	.31	1.24
	Mean latencies (in sec.)					
Exp. VII	11.10		13.75		14.56	14.72
Exp. VIII	10.50	12.64		13.40	13.95	

the center to the edge. As it appeared that this background was at a critical point in the density function, it was replaced with two other backgrounds, 2A and 2B, in which the average density was more strictly controlled and the distribution of dot elements followed the same plan as that of the denser backgrounds. Backgrounds 1 and 3 were retained but 4 was omitted.

Targets 2 and 3 were retained and two smaller targets, 4 and 5, were added. A different sample of 24 *S*s was tested by the same general procedure.

Discussion

The results were generally consistent with those of the previous experiment. The main effect for background density was significant, F (3, 69) = 9.441, $p < .001$, but post hoc comparisons again revealed that the significant differences were all between no background and all other background conditions. The apparent trend does suggest that the use of densities between those of 1 and 2A would make curve-fitting procedures appropriate, but it is doubtful if such backgrounds could be regarded as simple quantitative variations of density as the particular dispersion of elements around the target would be likely to become very important.

TABLE 7

TARGET SIZES AND MEAN LATENCIES

Display	Target diameters (in in.)				
	1	2	3	4	5
On glass slide	.91	.44	.22	.11	.05
On screen at start of trial	1.34	.65	.32	.16	.08
On screen at end of trial	4.11	1.98	.99	.50	.25
	Mean latencies (in sec.)				
Exp. VII	13.95	13.57	13.07		
Exp. VIII		13.14	12.60	12.19	12.55

The trend for target size was generally consistent with that obtained in Exp. VII, but due possibly to factors resulting from the poor optical resolution of the smaller targets the main effect was not significant. This would also account for the apparent reversal of trend for the smallest target.

It is probable that the use of more precise measures and a better optical system would lead to the detection of quantitative changes due to variation of both target size and density, but it seems clear that in both cases the effects are slight. The results appear to indicate that density variation has little effect on the detection of drift, but the qualitative change resulting from the introduction of a background of any density is the important factor.

EVIDENCE SUPPORTING THE TARGET-DRIFT THEORY: CANCELLATION OF TARGET DRIFT

Experiment IX: Qualitative Stimulus Variation and Cancellation of Drift

It has been shown that drift detection satisfies the criterion of being a function of the relevant stimulus information, but the criterion of accuracy is more difficult to

apply. This means that it is also difficult to compare the accuracy of this task with that of judgments of the location of X. It might also be argued that the demonstration of *O*'s ability to detect drift is not necessarily evidence that he can make use of this drift information. A task in which *S* actively utilizes the drift information and cancels the drift by moving the target in the opposite direction was therefore substituted for the passive detection of drift. The physics (but not the psychology) of this drift cancellation involve the movement of X into coincidence with the target, and the accuracy of performance can be assessed by measuring the discrepancy between the adjusted position of X and this position of coincidence. This measure would be directly comparable to the measure of accuracy in the expansion-information task—the discrepancy between the actual and judged locations of X.

The new task was made possible by the use of new apparatus in which the point source was mounted on a carriage and could be moved along a set of rails which ran on a horizontal path parallel to the screen. The *S* could control the position of the light (and hence of X) by means of a switch which made the light move to the left or right at a rate of about 15 mm/sec. The final position of the light could be read off to the nearest .5 mm.

The new apparatus differed in detail from that used previously but was of the same principle. The *S*'s position could be adjusted so that his eyes were aligned with the center of the screen. Head movements were gently restrained and extraneous noises reduced by means of a headrest consisting of an inverted pair of ear muffs. Viewing was binocular with no field stop.

The aim of Exp. IX was to demonstrate that *S*s could make use of detected drift, canceling it in a manner analogous to steering toward a target. It was also intended to show that although the accuracy of the task was affected by variation of the stimulus conditions, (Criterion 2) the effects were not extreme and the errors remained of an acceptable size (Criterion 1). The major stimulus variable was the nature of the expanding background to the target, the four backgrounds being chosen to represent a wide range of stimulus conditions. The backgrounds were (*a*) completely blank, (*b*) a rectangle with side projections as used in Exp. V, (*c*) an overall irregular dot pattern, density 25 dots per square inch, and (*d*) an overall fine-line grid as used in Exp. VI. The target, a disc 9 mm. in diameter, cast a shadow which varied from 11 mm. in diameter at the start of the trial to 39 mm. at the end.

The location of the "on target" position (zero heading error) was varied, as was the direction of the initial drift. One location was at he center of the screen and the other was 5 cm. from it at the same level and to *S*'s right. The initial heading errors were 5 cm. to the left and 5 cm. to the right of each correct location, producing drift to the right and left, respectively.

Method

Each *S* was given all conditions twice (a preliminary check showed that no change in accuracy occurred over 32 trials). The *S* was allocated a random permutation of the four background conditions and two random permutations of the four location and heading error combinations for each background. The 12 *S*s were each given a demonstration trial which used the background scheduled for the first block of eight test trials and were shown how to cancel drift by moving the target in the opposite direction. This was followed by four practice trials using a random permutation of two "on target" locations and two initial heading errors. The locations were 2 cm. to the left of center and 8 cm. to the right, and the initial heading errors were 4 cm. to the left or right of each location, producing drift to the right and left, respectively. Obvious faults of understanding were corrected and the 32 test trials administered.

Results

The absolute displacements of the light source from the "on target" position were taken as the error scores and subjected to an analysis of variance. The main effects of background and direction of initial drift were both significant, F (3, 33) = 13.48, $p < .001$, and F (1, 11) = 6.73, $.025 > p > .001$, respectively. The double interactions between background and location and location and direction of initial drift

TABLE 8

MEAN ERRORS IN MILLIMETERS

Location/drift cond.	Background			
	None	Rectangle	Dot	Grid
Center/left	6.81	7.06	12.21	12.77
Center/right	7.50	7.42	13.56	12.00
Off-center/left	3.75	4.44	9.35	12.27
Off-center/right	10.90	13.94	15.63	24.31

were also significant, F (3, 33) = 4.27, $.025 > p > .001$ and F (1, 11) = 27.40, $p < .001$. The means for each condition are presented in Table 8.

An examination of the mean errors shows that the aims of the experiment were achieved as the errors were small and remained so despite the significant effects of the experimental conditions. The off-center location (5 cm. to the right) produced mean errors for drift to the right that were larger than those for the other conditions, but even so the maximum error was less than 25 mm.

Discussion

The effects of the experimental conditions were examined separately because of the significant double interactions. For the central location, the direction of drift had a negligible effect. The effects of the backgrounds differed, with the mean errors for the no-background and rectangle conditions being very similar, with a sharp rise in mean error to the two similar means for the dot and grid backgrounds. This suggests a dichotomy between the effects of the two overall backgrounds, dot and grid, and those of the limited displays, rectangle and target alone. There appears to be a sudden increase in difficulty when an overall background is introduced, and for these results it may be more correct to regard the rectangle as merely an extension of the target rather than as a background.

The results for the off-center location were very different. The mean errors for drift to the left were generally the smallest for any location/direction condition, and an examination of the responses showed that they were atypical in that they tended to be overcorrections of drift. The mean errors for drift to the right were by far the largest, despite the fact that the target would have approached a stable reference frame (the edge of the screen) more closely in this condition than in any other. The differences between the two drift directions for this off-center location and between the direction results for the two locations, center and off center, cannot be explained as being due to central versus peripheral retinal stimulation, as *S*s would be fixating the target under all conditions.

The variation between the background means for the off-center location, particularly those for the drift to the right, suggests a consistent increase of mean error as the complexity of the background is increased. Thus the addition of the rectangle to the target provided more complexity, the overall dot pattern still more, and the highly integrated grid pattern the greatest degree of complexity of background.

These differences between the results for the central and off-center locations suggest that steering toward a target which is not centered in relation to *O* (because of a heading aspect required to counteract the effects of a cross-wind, for example,) presents a rather different situation to that in which *O* can feel that he is going "straight ahead" toward his desired target.

EVIDENCE SUPPORTING THE TARGET-DRIFT THEORY: COMPARISON OF CANCELLATION OF DRIFT PERFORMANCE WITH JUDGMENTS BASED ON EXPANSION PATTERN INFORMATION

Experiment X: Use of Expansion Information under Optimal Conditions

The errors obtained in Exp. IX may be compared with errors in judgments based on expansion information. As the conditions for maximum accuracy differ for the two tasks (for example, the most accurate condition for drift detection—no background—can hardly be used for the rival task), it was decided to use background and other conditions which would bias the results in favor of expansion-information judgments. The one background was used, the irregular small dot pattern having 3.25 dots per square inch as used in Exp. VIII.

This pattern provided a large number of elements covering the total area of the screen, but at the same time the dots would be far enough apart to be quite distinct. This basic pattern was modified by the addition of 12 larger circular discs (identical to the target in Exp. IX) which were scattered irregularly across the glass sheet. These would presumably help to define the expansion pattern and would provide further shape transformations. Conditions should therefore be appropriate for a Gibson-type judgment as well as for other possible types of judgment.

Although the pattern covered the whole screen, it would be dispersed enough to provide favorable conditions for a judgment of the Calvert type. The larger elements would provide extra "noticeable features" and potential fixation points and there would be multiple streamers and ample parafoveal stimulation. In case *S*s failed to make use of this availability of fixation points as a result of past experience, the experiment included a second stage in which a particular large element was specified as the target and fixation instructions given.

The display provided for the possibility that *O* seeks an element that is not moving or is moving only very slowly. When X was located 5 cm. to the right of the center of the screen, one of the large elements was only 2 cm. from it at the start of the trial and 6 cm. from it at the end. If the movement judgment depends upon relative lack of movement, there were two other large elements nearby that moved in different directions away from X, one moving from a position 9 cm. from X to one 29 cm. from it and the other moving from a position 9.7 cm. from X to one 31 cm. from it.

One possible use is very similar to Calvert's but does not have the restrictions of fixation and parafoveal stimulation. This is that *O* "notes the direction of motion of objects on the ground and interpolates back" to X (Palmer, personal communication, July 1970). While the off-center position provides for this the other location of X, the central position, might be more favorable. For this position, X lies toward the center of a triangle formed by the three large elements mentioned above. These elements were all within 8 cm. of X at the start of the trial, the closest being 5 cm. away from it. One moved straight up the screen, the second moved to *S*'s right and up at an angle of $7\frac{1}{2}°$, and the third moved to his left and down at an angle of 57°.

The instructions were designed to allow *S* the greatest freedom in method of judgment. To aid in this, the experiment was divided into two stages. For Stage 1, the instructions suggested that *S* might have a subjective impression that he was "heading" toward a particular point on the screen but might not have any idea why he had that impression. He was instructed to indicate that point. He was also instructed that if he should fail to gain this immediate awareness of where he was going, he was to try to work it out. To help in this, he was given instruction in the physical properties of the expansion pattern, using the diagrams from Exp. IV. For Stage 2, he was instructed to make the judgment while watching a particular large element, thus satisfying Calvert's requirements.

The comparison of the drift cancellation errors with the errors obtained in this experiment was biased in favor of the expansion-information theories because: (*a*) the criterion score from Exp. X was based on one background, designed to be very favorable to such judgments, while that from Exp. IX was based on overall performance on four backgrounds, some quite unfavorable to drift detection; (*b*) the locations of X, central and 5 cm. off center, would be most favorable, as errors in central locations have been shown to be at a minimum (at least partly because of artifactual factors); (*c*) the same *S*s were used with a delay of several months, so that short-lived effects such as fatigue would not carry over but sophistication should do so; (*d*) the length of the experimental session, under half an hour, should produce less boredom and fatigue than the sessions used in Exp. IX (about $1\frac{1}{4}$ hr.); and (*e*) limitations in the method of control of the light and setting of the stimulus

display would mean that errors obtained in Exp. IX would not be at their minimum value.

Method

The 12 *S*s from Exp. IX were given the instructions for Stage 1 and then presented with a random sequence of five trials on each location of X. The *S* indicated the judged position at the end of each trial. When Stage 1 was complete, the Stage 2 instructions were given, followed by five trials in which X was always at the center of the screen. The target specified in Stage 2 moved from a position 5 cm. from X to one 16.5 cm. from it, moving down and to the left at an angle of 57° to the horizontal.

Results and Discussion

The discrepancies between actual and judged locations of X were measured in millimeters. The *S* and group means for the two locations used in Stage 1, for Stage 2, and for Exp. IX over all conditions are presented in Table 9. The accurate results of the first *S* for the central position in Stage 1 probably reflect an unusually consistent response to the center of the screen, as his response to the off-center position had a mean error of 41.8 mm. while their mean distance from the center was only 17.5 mm.

TABLE 9

COMPARISON OF SUBJECT MEAN ERRORS (IN MILLIMETERS) FOR EXPERIMENTS IX AND X

S	Experiment and cond.			
	Exp. X			Exp. IX
	Stage 1, center	Stage 1, off center	Stage 2, center	Over all conditions
1	6.0	41.8	27.4	10.6
2	124.4	249.8	46.6	7.7
3	16.0	23.4	17.0	9.3
4	69.8	37.2	71.2	9.1
5	98.2	94.2	86.6	7.1
6	218.6	284.6	303.8	8.8
7	118.4	111.2	63.0	14.0
8	121.0	119.4	153.4	12.3
9	50.0	66.6	36.8	18.0
10	48.0	49.6	60.4	9.0
11	59.2	36.4	48.2	11.7
12	18.0	23.4	57.6	13.0
$\bar{X}$	78.97	94.8	81.0	10.9

Despite all efforts to bias the results in favor of expansion-information judgments, the errors were far in excess of those made in the cancellation of drift and the means (of 3–4 in.) were consistent with the mean errors made for such positions in the earlier experiments.

General Discussion

One limitation of all the studies involving use of the expansion pattern was that *S* always passively judged the position of X (or the "point being directly approached") and was never required to engage in active control of his "approach" to the display. It does not seem likely that *S* could do this active task when he could not locate X, and the type of display used would not allow one to distinguish between adjustments based on the information from the expansion and those in which *S* consciously or unconsciously made use of a target. Evidence from a display of a different type (a night-driving simulator currently being used at the University of New South Wales) suggests that adjustments based on the flow pattern are of limited accuracy. The *S*s commonly take up a path parallel to the "path" of the simulated taillights they are supposed to track, despite the presence of texture flows (up to a simulated 60 mph) which should reveal that the taillights do not line up with X.

The results of the driving simulator fit the drift-cancellation hypothesis as the position of the taillights was stabilized and drift canceled. In this case in which the target is itself locomoting and is not a part of the flowing surface, the cancellation of the drift merely ensures that the two paths of locomotion do not converge or diverge. To adjust the pursuit path so that it tracks directly behind the target, *O* has to make use of other information, for example, the angular orientation of the target, and this information is lacking in the simulator.

The expansion pattern is probably the basis (at least partly) of one's impression of self-movement during locomotion, but the particular displays and situations used gave the impression that the pattern was approaching. This illusory forward movement of the pattern proved a problem in the early experiments as it sometimes caused difficulty in focusing. The unpleasant visual effects did not occur when *S* was allowed unrestricted binocular vision of the screen, but his perceptual experience was still generally that of an integrated surface moving toward him. The presence of the

edges of the screen in his field of view and the use of binocular information were evidently sufficient to cause the effect to join the paradoxical ranks of television and movie displays, in which movement can be seen to be towards the viewer but at the same time limited to the surface of the screen (cf. Gregory, 1970). The determination of the conditions necessary to convert the impression of movement of the pattern into that of one's own forward movement remains one of the interesting problems in this field.

The illusory movement appears to be related to the role of the expansion pattern in maintaining size constancy in objects whose distance from *O* is changing. The expansion pattern can, under sufficiently reduced conditions, be made to yield an experience of what appears to be almost perfect size constancy (no apparent increase in element size or dispersion) at the expense of what might be called "distance constancy." This is the reverse of the usual effect in the traditional size constancy experiments which make use of static displays and appear to produce distance constancy at the expense of size constancy.

Other potential roles of the expansion pattern include the specification of the three-dimensional modeling of the approached surface (e.g., concave, convex, and bumpy) and the production of the integrated, unified surface that is experienced when it might be more reasonable to expect the impression of a swarm of dots. The evidence that the expansion pattern is not responsible for the guidance of locomotion should, therefore, not be taken as evidence that it is not of importance in perception.

REFERENCES

CALVERT, E. S. Visual aids for landing in bad visibility, with particular reference to the transition from instrument to visual flight. *Transactions of the Illuminating Engineering Society, London*, 1950, **15**, 183–219.

CALVERT, E. S. Visual judgments in motion. *Journal of the Institute of Navigation, London*, 1954, **7**, 233–251.

GIBSON, J. J. *The perception of the visual world*. Boston: Houghton-Mifflin, 1950.

GIBSON, J. J. The visual perception of objective motion and subjective movement. *Psychological Review*, 1954, **61**, 304–314.

GIBSON, J. J. Optical motions and transformations as stimuli for visual perception. *Psychological Review*, 1957, **64**, 288–295.

GIBSON, J. J. *The senses considered as perceptual systems*. London: Allen & Unwin, 1966.

GIBSON, J. J., & GIBSON, E. J. Continuous perspective transformations and the perception of rigid motion. *Journal of Experimental Psychology*, 1957 **54**, 129–138.

GREGORY, R. J. *The intelligent eye*. London: Weidenfield & Nicolson, 1970.

PALMER, E. A. Determination of human ability to perceive aircraft aim point from expanding gradient cues. Paper presented at the meeting of the Aerospace Medical Association, May 1969.

(Received April 30, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 262–267

IMPLICIT SPEECH INFERRED FROM RESPONSE LATENCIES IN SAME–DIFFERENT DECISIONS [1]

STUART T. KLAPP [2]

California State College at Hayward

When *S*s read numbers or words aloud, the latency from stimulus onset to the initial vocalization depends upon the number of syllables to be pronounced. The present experiments showed that response latency also depends on syllables for same–different decisions, thereby suggesting that the effect of syllables operates during stimulus recognition rather than during preparation for overt vocalization. The interpretation was that a central process, preparatory for and more rapid than overt speech, is involved in comprehension of written language symbols.

Theories of language comprehension which involve proprioceptive feedback from articulatory responses may be rejected since comprehension seems to be unaffected by total paralysis of the articulatory organs (Smith, Brown, Toman, & Goodman, 1947) and since reading with comprehension can occur at a more rapid rate than overt word articulation. However, such considerations do not rule out a theory in which internal or implicit speech rather than motor representation of speech is involved in language comprehension. This theory has had the undesirable property that few clear predictions have been generated by it. However, Eriksen, Pollack, and Montague (1970) have suggested that reaction time in tasks requiring comprehension of written material should increase as a function of the time that would have been required to articulate the stimulus. They reported a series of experiments in which *S*'s task was to read single words or two-digit numbers aloud as soon after visual presentation as possible. The latency from the onset of the printed word or number until the beginning of overt vocalization was found to increase as the number of syllables to be pronounced increased, leading Eriksen et al. to conclude that *S*s "implicitly speak a word before overtly pronouncing the word."

These data, however, do not necessarily suggest that implicit speech is needed for comprehension of the word or number. Instead, the results might indicate that an implicit speech process is required only to set up the vocal apparatus for production of overt speech, while comprehension itself occurs with a latency independent of the number of syllables. If this is the case, then the increase in response latency with the number of syllables should be observed only if *S* must pronounce the stimulus and not if *S* were to indicate recognition without pronouncing the stimulus itself. On the other hand, if implicit speech is involved in the comprehension of the stimulus, then the response latency should depend on the number of syllables even if *S*'s task does not require pronouncing the stimulus. An example of such a task would be deciding whether two stimuli are the same or different. Experiment I was an attempt to replicate the Eriksen et al. (1970) findings and then to determine if the increase in latency with syllables would be found for a same–different decision task as well as an overt naming task.

EXPERIMENT I

Two-digit numbers requiring two, three, or four syllables to pronounce were presented to two independent groups of *S*s, who were to read the numbers aloud (R

[1] This research was supported, in part, by Institutional Grant Funds for the Advancement of Science from the National Science Foundation. The author wishes to express his appreciation to Gary Rapp, who assisted in constructing the apparatus and in gathering the data of Exp. I, and to Charles G. Burgett, who gathered the data of Exp. III. Portions of these data were presented at the meetings of the Western Psychological Association, San Francisco, April 1971.

[2] Requests for reprints should be sent to Stuart T. Klapp, Department of Psychology, California State College, Hayward, California 94542.

condition) or to determine whether two, two-digit numbers were the same or different (S/D condition). Experimental trials in which a response was required immediately upon stimulus presentation (IMM trials) were compared to control trials in which the stimulus was viewed for 2 sec. prior to a signal to respond (SIG trials). This control was used to assure that any effect of syllables upon response latency could be attributed to the processing of the stimulus rather than to artifacts of the way particular sound patterns activated the timing apparatus. The SIG control condition was used by Eriksen et al. (1970) for word stimuli but not for number stimuli, which were the only stimuli in which syllable length and length of visual stimulus were not confounded. Therefore, on technical grounds, a replication using this control with the number stimuli seemed desirable. Thus Exp. I had a $3 \times 2 \times 2$ factorial design, with 3 within-*S* conditions of syllable length, IMM and SIG conditions also varied within *S*s, and the R and S/D tasks, which were given to independent groups of *S*s. Four daily sessions of 108 trials each were given in order to observe any interaction of the syllable effect with session-by-session practice.

Method

Subjects.—The *S*s for all three experiments were students in introductory psychology at California State College, Hayward, who participated in order to fulfill a course requirement. They were told that the purpose of the experiment was to "study reaction time in reading for normal adults" and were permitted to examine the apparatus in order to alleviate any anxiety. Alternate *S*s were assigned to the R and S/D conditions of Exp. I in the order in which they reported for the experiment.

Materials.—The stimuli were black two-digit numerals subtending a visual angle of 1.6° high × 2.4° wide. The two-syllable numbers were 14, 15, 20, 30, 40, 50, 60, 80, and 90; the three-syllable numbers were 17, 28, 39, 44, 52, 61, 70, 83, and 95; and the four-syllable numbers were 27, 37, 47, 57, 67, 76, 78, 87, and 97. These are the numbers used by Eriksen et al. (1970). and they appear to be a representative sample of the 13 possible two-syllable, 59 three-syllable, and 15 four-syllable numbers. For the S/D condition, the second two-digit number was approximately 7.4° to the right and 2.6° below the first number. For each *S* in the S/D condition, a randomly selected half of the trials involved the same numbers, while the remaining trials had different numbers of the same syllable length.

Apparatus.—The *S* was seated in a dimly illuminated Industrial Acoustics sound isolation chamber, approximately 68 cm. from a rear projection screen on which the stimulus numbers were projected with a Kodak Carousel projector. This projector, together with the experimental equipment and *E*, was located in a second isolation chamber. The *S*'s latency was measured by a Hunter (Model 120) Klockounter, stopped by the sound of *S*'s voice as picked up by an Electro-Voice (676) microphone, amplified, and then fed into a Gerbrands voice relay. Although *S* was encouraged to speak at a normal level, the sensitivity of this equipment was sufficient to stop the timer with a slight whisper. For IMM trials this timer was started by a photoelectric cell activated by light reflected from the screen when the shutter in front of the projector was opened. For SIG trials, the timer was started by a relay which simultaneously activated a neon signal light in *S*'s room. The events making up each trial were sequenced and timed automatically by a motor-driven switch with a cycle between successive stimulus presentations lasting 6 sec.

Procedure.—Each of the 24 *S*s received four daily sessions consisting of four blocks of 27 trials, of which 9 trials involved two-syllable numbers, 9 three syllables, and 9 four syllables. The order in which these numbers were presented was randomized for at least every two *S*s, producing a random pattern of two-, three-, and four-syllable numbers in each block of trials. For half of the *S*s, each daily session contained a block of 27 IMM trials, followed by two blocks of 27 SIG trials and then one block of 27 IMM trials. For the remaining *S*s, the blocks were given in the complementary order. All *S*s received 3 practice trials prior to the first block of IMM trials and the first block of SIG trials.

The *S*s in the R condition were instructed to "read the number aloud (e.g., 13 is read as 'thirteen') as quickly as possible" for IMM tirals. For the SIG trials the instructions were to "look at the number when it appears on the screen but remain silent until the signal light comes on 2 sec. later" and then to "say the numbers as quickly as possible." For *S*s in the S/D condition, similar instructions were used except that *S* was to say "yes" if the two numbers were the same and "no" if they were different. The responses of "yes" and "no" were chosen so that either response would be a monosyllable. The *S*s were informed that reaction time was the variable under study in the experiment.

In the event that an erroneous response was made or stray noise (coughing, chair movements, etc.) activated the equipment, the trials were repeated at end of the daily session. No feedback concerning either latencies or errors was provided.

Results and Discussion

The overall rate of incorrect responses was .68%, with 1.28% in the S/D–IMM

TABLE 1

Response Latencies for Experiment I

Cond.	Syllables to pronounce			
	2	3	4	$\bar{X}$
R–IMM	503	522	536	520
R–SIG	391	398	391	393
S/D–IMM	645	662	678	662
S/D–SIG	341	337	340	339

Note.—Latencies are given in milliseconds.

condition; .68% in R–IMM; .54% in R–SIG; and .22% in S/D–SIG. The total rate of unusable trials, including those on which stray noise activated the apparatus as well as incorrect responses, was 1.57%. All of these trials were repeated and the data included in the analyses.

Since there was no consistent interaction of sessions of practice by the effect of syllables, the data were collapsed across the four daily sessions for the analyses to be reported. The mean latencies are given in Table 1. Overall, the SIG conditions yielded substantially shorter latencies compared to the IMM conditions, $F\ (1,\ 22) = 230$, $p < .001$, a finding consistent with the premise that analysis of the stimulus would occur prior to the response signal in the SIG trials. This is the premise on which the use of this condition as a control is based. The main effect of the R versus S/D conditions on response latency was nonsignificant, $F\ (1,\ 22) = 2.1$, $p > .05$; however, there was a significant interaction, $F\ (1,\ 22) = 43$, $p < .001$, such that the S/D condition had longer latency than the *R* condition for IMM trials with the reverse relationship for the SIG trials. This aspect of the data is not directly relevant to the present purposes.

The main effect of the number of syllables on response latency was significant, $F\ (2,\ 44) = 23$, $p < .001$, with an interaction, $F\ (2,\ 44) = 31$, $p < .001$, such that the effect of syllables was more pronounced for IMM trials than for the SIG control trials. This main effect and interaction represent the phenomenon of primary interest in this investigation. The lack of a significant three-way interaction, $F\ (2,\ 44) < 1$, indicates that the critical two-way interaction of syllable effect by IMM versus SIG does not differ significantly for the R condition as compared to the S/D condition. The nonsignificant two-way interaction of syllables by R versus S/D, $F\ (2,\ 44) = 1.07$, $p > .10$, also is consistent with the conclusion that the syllable effect does not depend upon the R versus S/D task difference. These findings can be more simply described and interpreted by examining the R and S/D conditions separately. Although statistical tests are reported for these conditions separately, such analysis was done subsequent to obtaining a significant overall F for the syllable effect and its interaction with IMM versus SIG control.

First consider the R condition, in which the IMM trials represent a replication of Exp. III of Eriksen et al. (1970). The overall effect of syllables on latency in the R–IMM condition was significant, $F\ (2,\ 22) = 20$, $p < .001$, and the slope of the regression line fitted to these data was 16.5 msec. per syllable, compared to 11 msec. reported by Eriksen et al. (1970), Exp. III. The larger estimate in the present experiment might be due to the lack of feedback on latency performance with a consequent reduction in motivation to keep latencies down. The present experiment, unlike the reference experiment, used the control SIG trials with the number stimuli. As expected, there was no apparent effect of syllables in the SIG trials, and the interaction of IMM versus SIG with the syllable effect was significant, $F\ (2,\ 22) = 14.7$, $p < .001$. These findings represent a replication and further confirmation of the effect reported by Eriksen et al. (1970).

Of particular interest in the present investigation was the effect of the number of syllables on latency in the S/D condition. For the S/D-IMM trials, the effect of syllables on response latency was significant, $F\ (2,\ 22) = 21$, $p < .001$, and the slope of the regression line fitted to these data was 17.0 msec. per syllable, which approximates the 16.5 msec. per syllable in the R–IMM condition. There was no apparent effect of syllables in the control SIG trials, and the interaction of IMM versus SIG with the syllable effect was significant, $F\ (2,\ 22) = 16.7$, $p < .001$.

The finding that response latency increases with the number of syllables even for the S/D condition in which no overt number-pronouncing response was required is the critical finding for the present investgiation. According to the argument presented earlier, this suggests that an implicit speech process is involved in the comprehension of the numbers.

The same and different response latencies of the S/D–IMM trials are considered separately in Table 2. Latencies for same responses were significantly faster than latencies for different responses, $F\ (1, 11) = 16, p < .01$, which is in agreement with the findings frequently reported (e.g., Krueger, 1970; Nickerson, 1965, 1968). The interaction of the syllable effect by the type of response was complex and significant, $F\ (2, 22) = 20, p < .001$. In view of the importance assigned to the effect of number of syllables in the S/D condition, the nonmonotonic nature of the functions for latencies when same and different responses are considered separately is disturbing. Therefore, a replication was attempted in Exp. II.

TABLE 2

Latencies for S/D–IMM, Experiment I, Broken Down by Type of Response

Response	Syllables to pronounce			
	2	3	4	$\bar{X}$
Same	626	670	652	649
Different	665	655	705	675
$\bar{X}$	645	662	678	

Note.—Latencies are given in milliseconds.

Experiment II

A replication of the S/D–IMM condition was attempted with a manual response to indicate same or different decisions rather than the vocal "yes" or "no" responses used in Exp. I. This change in response modality was made to determine if the basic finding suggesting implicit speech in number comprehension could be generalized to a task in which *S* was not required to make any overt vocal response whatsoever.

Method

The details of the procedure were the same as for the corresponding condition of Exp. I, except that *Ss* indicated their response by use of a switch held in the preferred hand. For half of the *Ss*, a leftward movement of the response switch indicated same, and for the remaining half of the *Ss* the position of the switch was reversed. This switch stopped the timer used to determine response latency, replacing the voice relay used in Exp. I. Since a manual response, unlike the verbal response, requires that new response habits be formed, each *S* was given one block of 27 unscored trials. Then each of the 12 *Ss* received four blocks of 27 trials, of which 9 involved numbers of each syllable length. Randomization was carried out as in Exp. I.

Results and discussion

The rate of errors was 1.70%, which was somewhat higher than the 1.28% incorrect responses in Exp. I, S/D–IMM condition. The additional errors in Exp. II may be attributed to incorrect motor responses which may occur even though *S* could have made a correct vocal response. The overall mean latency in Exp. II was 667 msec., only slightly longer than the mean latency of 659 msec. for the S/D–IMM condition of Exp. I. The data of Exp. II are presented in Table 3, which may be compared to the corresponding data of Exp. I. in Table 2.

In agreement with the corresponding S/D–IMM condition of Exp. I, the latency, averaged across same and different responses, increased with syllables, $F\ (2, 22) = 13, p < .001$, with a regression of 20.2 msec. per syllable which approximated the 17.0 msec. per syllable for the corresponding condition of Exp. I. Thus, the finding that response latency increases with the number of syllables which would have been needed to pronounce the number has been replicated for the condition in which *S* indicated his same or different decision manually rather than vocally. Also in

TABLE 3

Response Latencies for Experiment II

Response	Syllables to pronounce			
	2	3	4	$\bar{X}$
Same	600	648	663	637
Different	699	675	717	697
$\bar{X}$	650	663	690	

Note.—Latencies are given in milliseconds.

agreement with Exp. I was the faster latency for same as opposed to different responses, F (1, 11) = 52.5, $p < .001$, and the significant interaction of the syllable effect by same versus different responses, F (2, 22 = 13.4, $p < .001$. Once again the details of the interaction were complex, so that still another replication seemed desirable.

Experiment III

The above two experiments, as well as Exp. III reported by Eriksen et al. (1970), used two-digit numbers requiring two, three, or four syllables to pronounce. Unfortunately, other properties of the numbers were unavoidably confounded with the number of syllables. Most of the two-syllable numbers contained the digit 0, while all of the four-syllable numbers contained the digit 7. It is possible that these confounded properties of the numbers, rather than syllables, may have produced the increase in latency. Although Eriksen et al. (1970) also reported latency data for words, their word stimuli always confounded the number of letters with the number of syllables. Moreover, they observed only latencies for pronunciation, and not for same–different decisions. Thus, a replication of the S/D–IMM condition, using words rather than numbers as stimuli, seemed desirable.

The response required of S was modified slightly to assure that the previous results were not some artifact of S's strategy when both same and different responses were required during the same block of trials. In Exp. III, S was to respond by pressing a switch only if the two words were the same, while making no response if the words were different for one entire block of trials, the first block for half of the Ss. Then for the other block of trials, responses were to be made only when the words were different.

Method

Materials.—The 28 words all had three consonants and two vowels and were chosen to form seven sets of four words, with two words of one syllable and two words of two syllables in each set. All four words of each set started with the same consonant and had the same frequency-of-occurrence classification in the Thorndike-Lorge count. The word pair combinations to be judged same or different were generated from these matched sets of four words such that all possible pairings of words having the same number of syllables were used. For example, the set of words consisting of clear, court, cover, and color yielded the following eight combinations: clear-clear, court-court, clear-court, court-clear, color-color, cover-cover, color-cover, cover-color. The other six sets of words used in this manner were: false, frame, fancy, final; force, field, favor, fifty; heard, horse, happy, honor; large, learn, labor, lower; paint, pound, paper, power; and taste, teach, table, taken. This method produced stimulus word pairs with the same initial letter, insuring that S could not base his response on a comparison of initial letters.

The 56 word pairs were photographed in capital black letters such that the first word of each pair was in the upper left and the second word was in the lower right. As projected for viewing by S, the letters were 1.1° high, with a horizontal offset between the first letters of the two words of 4.5° and a vertical offset of 6.3°.

Apparatus.—The S responded with a single push-button held in his preferred hand. Except for this change, the apparatus was the same as that used in Exp. II.

Procedure.—Each of the 12 Ss was given two blocks of 56 trials, each of which used all 56 word pairs. For one entire block of trials, S was instructed to press the response button when the two words were the same and to make no response when the words were different. For another block of 56 trials, a response was to be made when the words were different. Alternate Ss were to make either same or different responses during the first block. The order in which the 56 combinations of words were presented was randomized for every two Ss. Instructions to S were to "respond as quickly as you can, while avoiding errors," and Ss were informed that reaction time was the variable of interest in the experiment. Errors of response comission were recorded, but the latencies of these responses were not considered. No errors of omission were anticipated since the stimuli remained in S's view for 3 sec.

Results

The Ss made 31 false positive responses, which represents 4.6% of the possibilities for making such errors. This is considerably higher than the rate of errors in Exp. I and II and probably reflects the difficulty of inhibiting responses rather than any difficulty in recognizing words. The average response latency was 695 msec., which is somewhat longer than the latencies of the first two experiments. These data are pre-

sented in Table 4, which may be compared to the corresponding data of Exp. I and II in Tables 2 and 3.

Consistent with the previous experiments, stimuli requiring two syllables rather than one to pronounce led to longer latencies of same–different decisions, $F(1, 11) = 10.82$, $p < .01$. The difference between mean latencies for two- and one-syllable words was 23 msec., which closely approximated the 20.1 msec. per syllable effect in Exp. II. Also, in complete agreement with the previous results were the longer latencies for different compared to same responses, $F(1, 11) = 10.16$, $p < .01$. The interaction of same versus different responses with the syllable effect was significant, $F(1, 11) = 9.55$, $p < .05$, and was such that the syllable effect was obtained for different responses, but not for same responses. Although the details of this interaction did not agree with the corresponding interaction in Exp. I and II, an effect of syllables was found in all three experiments.

TABLE 4

RESPONSE LATENCIES FOR EXPERIMENT III

Response	Syllables to pronounce		
	1	2	$\bar{X}$
Same	662	662	662
Different	705	751	728
$\bar{X}$	684	707	

Note.—Latencies are given in milliseconds.

General Discussion

A consistent and significant main effect of syllables on same–different judgment latency was observed regardless of whether words or numbers were the stimuli and regardless of the details of the nature of the response required of *S*. The magnitude of the increase in latency, approximately 20 msec. per syllable, is much shorter than the time required for overt speech, which is approximately 100 msec. per syllable in the data reported by Landauer (1962). Indeed, the rate of processing is quite fast enough to permit silent reading at normal rates. Lewis (1958) reported that adults can read at about 600 words per minute after brief training in speed reading. Assuming an average of two syllables per word, this rate of reading implies a processing time of 50 msec. per syllable. Apparently the rate of silent reading is limited by factors other than the rate of processing observed under the conditions of the present experiments. On the other hand, far faster rates of reading are sometimes reported under highly specialized conditions of instruction. Such rates might be achieved at the expense of full processing of each individual word in the text.

These results can be taken to suggest that an implicit representation of speech is involved in comprehension of printed words and numbers. Since the rate at which this implicit speech occurs is much faster than overt articulatory movements, a peripheralist interpretation does not appear plausible. Rather, the results suggest that a central "preprogramming" for speech processes may be involved in comprehension. Of course, it is almost certainly the case that some tasks involving printed numbers or letters can be accomplished without implicit speech, e.g., locating the numeral 8 in a page full of 7's would not necessitate this degree of processing of each individual 7. And clearly implicit speech can hardly underly recognition of nonspeech stimuli, for which other forms of implicit responses may be involved (Norton & Stark, 1971). But for tasks involving comprehension of written language symbols, the data seem to suggest that high-speed implicit speech is involved.

REFERENCES

ERIKSEN, C., POLLACK, M., & MONTAGUE, W. Implicit speech: Mechanism in perceptual encoding? *Journal of Experimental Psychology*, 1970, **84**, 502–507.

KRUEGER, L. Effect of bracketing lines on speed of "same"—"different" judgments of two adjacent letters. *Journal of Experimental Psychology*, 1970, **84**, 324–330.

LANDAUER, T. K. Rate of implicit speech. *Perceptual & Motor Skills*, 1962, **15**, 646.

LEWIS, N. *How to read better and faster* (3rd ed.) New York: Thomas Crowell, 1958.

NICKERSON, R. Response times for "same"—"different" judgments. *Perceptual & Motor Skills*, 1965, **20**, 15–18.

NICKERSON, R. Note on "same"—"different" response times. *Perceptual & Motor Skills*, 1968, **27**, 565–566.

NORTON, D., & STARK, L. Scanpaths in eye movements during pattern perception. *Science*, 1971, **171**, 308–311.

SMITH, S., BROWN, H., TOMAN, J., & GOODMAN, L. The lack of cerebral effects of D-Tubocurarine. *Anesthesiology*, 1947, **8**, 1–14.

(Received May 10, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 268–272

MIRROR-IMAGE CONFUSABILITY IN ADULTS [1]

PETER WOLFF [2]

University of Wisconsin

Several studies have indicated that children have difficulty differentiating mirror-image stimuli. In the present study adults were required to classify pairs of horseshoe stimuli as same or different. Response times were compared for stimulus pairs that varied in orientation (left–right vs. up–down) and spatial plane of the pair (horizontal vs. vertical). Stimulus pairs in which the orientation matched the spatial plane of the pair (i.e., horizontal and left–right or vertical and up–down) took longer to classify than stimulus pairs in which these two variables were crossed. These results are interpreted as reflecting the necessity of synthesizing two sources of information in order to compare the former pair types—temporally encoded visual information and directional information from the motor scanning process. Implications for the source of children's difficulty with mirror-image stimuli of this type are discussed.

The discrimination of mirror-image stimuli is known to be uniquely difficult for children (Huttenlocher, 1967a, 1967b; Rudel & Teuber, 1963; Sekuler & Rosenblith, 1964), as well as many species of lower animals (Mackintosh & Sutherland, 1963; Riopelle, Itoigawa, Rahm, & Draper, 1964; Sutherland, 1957). Using three-sided horseshoe-shaped stimuli in a horizontal plane, Rudel and Teuber found that children confused left–right oriented horseshoes (i.e., $\subset \supset$) but not up–down oriented ones (i.e., $\cap \cup$). While these findings led investigators to believe that the phenomenon of mirror-image confusability was specific to left–right oriented stimuli, subsequent research demonstrated that the relative position of the pair of mirror-image stimuli is important in determining their discriminability. When the stimuli were in the horizontal plane, left–right oriented horseshoes were confused more often than up–down pairs, as previous studies had shown. However, when the pairs were in the vertical plane (i.e., $\substack{\cup \\ \cap}$ or $\substack{\supset \\ \subset}$) the reverse was true—confusability was greater for up–down than left–right oriented stimuli (Huttenlocher, 1967a, 1967b; Sekuler & Rosenblith, 1964).

Huttenlocher (1967a) suggested that the important factor in mirror-image confusability is the orientation of the stimulus pair in relation to its axis of separation. In her study, nursery school children were required to match the orientation of a standard horseshoe. Each child's horseshoe was to be placed either beside or below the standard. She found that children made errors in their placement only when rotation of a correctly placed test horseshoe around the axis separating the pair would change its orientation relative to the standard. In other words, errors were made on this type of configuration, $\subset\!\vdots\,$X but not this, $\cap\!\vdots\,$X, where "X" denotes the position of the test horseshoe and the dotted line represents the axis of separation of the pair. Huttenlocher called the former pair type "mirror-image," and the latter pair-type "aligned." Her data also indicate that more errors were made on left–right than up–down oriented stimuli and among mirror-image stimuli, more errors on pairs in the horizontal plane than in the vertical plane.

While the existence of mirror-image confusability in children and animals is un-

[1] This research was conducted while the author was at the Wisconsin Research and Development Center for Cognitive Learning, which is supported by grants from the United States Office of Education, Department of Health, Education, and Welfare under the provisions of the Cooperative Research Program (Center No. C-03, Contract OE 5-10-154).

[2] Requests for reprints should be sent to Peter Wolff, Wisconsin Research and Development Center for Cognitive Learning, 1404 Regent Street, Madison, Wisconsin 53706.

questioned, there is little agreement about a suitable explanation for the phenomenon. Caldwell and Hall (1969, 1970), have recently claimed that the child's difficulty with mirror-image stimulus pairs in a matching-to-sample task results from the child's inadequate definition of the concepts "same" and "different" as applied to mirror images. Since prereading children have always regarded mirror images of an object as the same object, this habit is carried over to the judgment of mirror-image letter pairs. These authors demonstrated significant improvement in discrimination of these pairs after a training task which they felt defined mirror-image stimuli as different. Gibson, Gibson, Pick, and Osser (1962) offer a similar explanation for the tendency of children in their study to confuse mirror-image letter-like pairs.

Corballis and Beale (1970), on the other hand, have proposed that mirror-image confusion arises from the bilaterally symmetrical organization of the organism's nervous system. They demonstrate that a perfectly bilaterally symmetrical system would be unable to distinguish left from right, and hypothesize that left–right confusability disappears in the child when handedness or consistent left–right scanning habits, made possible by the developing asymmetry of the brain, become established.

While Caldwell and Hall's (1969) explanation has the advantage of simplicity, it violates the fact that adults are known to confuse mirror-image stimuli, as, for example, in writing d for b or d for g, and direction, as in the case of a driver who signals a left turn and then turns right. There is, however, almost no empirical evidence concerning the discrimination of mirror-image stimuli for adults.

The present study examined the relative difficulty of mirror-image and aligned-stimulus pairs, using adult *S*s but following as closely as possible the stimulus material and procedures used by Huttenlocher (1967a) with children. Time to respond "same" or "different" to a pair of horseshoe stimuli was used as a measure of difficulty since performance by adults in Huttenlocher's task would presumably be error free. Applying Huttenlocher's terminology, a mirror-image pair is one for which the rotation of one of the horseshoes around the axis of separation of the pair changes the identity status of the pair from either same to different (i.e., ⊂ ⊂) or different to same (i.e., ⊂ ⊃). An aligned pair is one for which this rotation does not change the identity status of the pair (i.e., ∪ ∪ or ∪ ∩).

If, as Caldwell and Hall (1969) propose, children's mirror-image confusions simply reflect a definitional problem arising from the tendency of the child to call different orientations of a figure the same figure, then adults who thoroughly understand the requirements of the task should compare mirror-image stimulus pairs as quickly as aligned pairs.

If the bilateral symmetry explanation of Corballis and Beale (1970) is a sufficient explanation of the phenomenon, any superiority of aligned over mirror-image pairs should be restricted to pairs located in the horizontal plane, since the human organism does not possess symmetry around the horizontal median plane.

Method

Subjects.—Twenty-four undergraduates with normal or corrected-to-normal vision served individually as *S*s.

Materials and appartaus.—Stimuli were 16 pairs of horseshoe-shaped figures. The horizontal set consisted of the following pairs:

⊂⊂, ⊃⊃, ∪∪, ∩∩, ⊂⊃, ⊃⊂, ∩∪, and ∪∩.

Pairs 1, 2, 5, and 6 are mirror-image pairs. The remaining pairs are aligned. The vertical set consisted of 90° rotations of each of these pairs. The horseshoe pairs were drawn with black India ink and the final stimuli were duplicated from these drawings by photo offset onto heavy white stock. The sides of the horseshoes were 3.2 cm. long (subtending 4.2° of visual angle) and 1.2 mm. wide (.30° of visual angle). The two horseshoes making up a stimulus pair were separated by 1.3 cm. (1.7° of visual angle).

Stimuli were presented in a Polymetric two-channel tachistoscope (Model V-0959) with a blank fixation field. The tachistoscope was wired with a Hunter Klockcounter and a response panel containing two buttons, one for a "same" response and the other for a "different" response. Exposure of a

stimulus started the clock, and depression of one of the response buttons stopped the clock and turned off the stimulus.

Procedure.—Before responding in the reaction time (RT) task, *S* was shown each of the stimulus pairs and required to say whether their orientation was the same or different. He was told that when a stimulus pair appeared in the tachistoscope he was to respond "same" or "different" by pressing the appropriate response button. The *S*s were instructed to respond as quickly as possible but to avoid making errors.

Each session was divided into two parts, separated by a 5-min. rest period, with the horizontal pairs presented in one part and the vertical pairs in the other. The *S*s responded to 11 replications of each series of eight stimuli. The first presentation of each stimulus served as a practice or warm-up trial and was not included in the analysis. Each replication was presented in a different random order. Presentation of each stimulus was preceded by approximately 1 sec. by a "ready" signal. Half of the *S*s received the horizontal pairs in the first part, while the remaining half received the vertical pairs. For half of each order group the "same" response was made with the preferred hand, while the remaining half used the nonpreferred hand. The correctness and latency of each response was recorded.

Results

Mean RTs for the various pair types are shown in Table 1. Aligned stimulus pairs were responded to 57 msec. faster than mirror-image pairs, $F\ (1,\ 20) = 49.23$, $p < .001$. This difference is less pronounced for "same" pairs (39 msec.) than for "different" pairs (75 msec.), $F\ (1, 20) = 15.21$, $p < .001$, but is highly significant for both ($F = 29.39$ and 43.68, respectively, $p < .001$).

For stimuli in the horizontal plane Aligned pairs were responded to 51 msec. faster than mirror-image pairs. For vertical pairs this difference was 62 msec. Separate analyses showed each of these differences to be highly significant, $t\ (23) = 5.45$ and 5.27, respectively, both ps $< .001$. The difference between these two values, which is equivalent in this stimulus set to the difference between up–down and left–right orientation, does not approach significance ($F < 1$).[3] While there was no difference in RT to up–down and left–right orientation, when the stimuli were the same up–down stimuli were responded to slightly faster than left–right stimuli. For different pairs, the reverse was true. This interaction is significant, $F\ (1, 20) = 7.21$, $p < .01$.

Responses to "same" pairs were 25 msec. faster than those to "different" pairs, $F\ (1, 20) = 17.85$ $p < .001$, although this effect must be evaluated with the significant Response × Pair Type interaction reported above. For mirror-image pairs, "same" responses were made 42 msec. faster than "different" responses. For aligned pairs, this difference was only 7 msec.

A significant Plane × Order of Presentation interaction was found, $F\ (1,\ 20) = 27.14$, $p < .001$, reflecting the fact that RT decreased from the first to the second half of the session.

While the error rate was only 3.7%, and the differences among conditions small, more errors were made on mirror-image than on aligned pairs and more on horizontal than on vertical pairs (see Table 1).

TABLE 1

MEAN RT (IN MSEC.) FOR ALIGNED AND MIRROR-IMAGE STIMULUS PAIRS IN THE HORIZONTAL AND VERTICAL PLANE

Cond.		Aligned	Mirror image
Horizontal	same	554 (18)	598 (22)
	different	573 (16)	631 (22)
	mean	564 (17)	615 (22)
Vertical	same	583 (14)	615 (15)
	different	577 (16)	668 (19)
	mean	580 (15)	642 (17)

Note.—Parentheses indicate number of total errors.

Discussion

The results of this study demonstrate conclusively that adults have difficulty comparing mirror-image stimulus pairs when "mirror image" is defined in terms of the axis of separation of the pair. Furthermore, this difficulty is found equally for pairs in the vertical and horizontal planes.

One study with adult *S*s using a same–different RT task and stimuli of this type has been reported by Sekuler and Houlihan (1968).

[3] Within the constraints of this stimulus set, the alignment factor (mirror image vs. aligned) is actually the interaction of relative placement of the pair (horizontal vs. vertical) and orientation of the stimuli (up–down vs. left–right). Its "main effect" status is thus theoretical and not statistical.

Translating their results into the language of this article, the differences between mirror-image and aligned stimuli were significant for horizontal, but not for vertical pairs. These results must be interpreted with caution, however, since these differences were numerically almost identical, 42 and 38 msec. for horizontal and vertical pairs, respectively. Their failure to find significant differences between mirror-image and aligned pairs in the vertical plane is probably due to several factors contributing to instability in their data. The *S*s were uncertain as to the plane in which the pairs would appear since presentation of horizontal and vertical pairs were intermixed. In addition, *S*s responded by moving a toggle switch to the right or left. A directional response of this type may be incompatible with a task requiring comparison of stimulus directionality. These factors, as well as their use of median values of only six responses to each stimulus, may explain the fact that their median RTs were almost 200 msec. longer than the mean RTs found in the present study.

The fact that adult response times are longer for mirror-image stimuli argues against the claim by Caldwell and Hall (1969, 1970) that children's errors are caused merely by their misunderstanding of the definition of identity applied to mirror-image pairs. In the present study, *S*s clearly understood the relevant distinction to be made before starting in the RT task. Also, since mirror-image difficulty was no greater for horizontal than for vertical pairs, and RT was the same for up–down and left–right stimuli, the bilaterally symmetrical organization of human adults cannot explain their performance on this task.

A plausible explanation of these findings, based on a suggestion by Deutsch (1955), depends on the fact that for mirror-image pairs the orientation of the individual stimuli (right–left or up–down) always matches the plane of the pair (horizontal or vertical). For aligned pairs the opposite is true—up–down stimuli are in the horizontal plane, while left–right stimuli are in the vertical plane. If it is assumed that *S* either explicitly or implicitly scans the pair of stimuli, either from the center outward in each direction, or starting at one end of the array, then in order for *S* to determine the orientation of a mirror-image stimulus he must integrate sensory information which is temporally organized with motor information about the direction of his scan. This point is most clearly seen in the example of a pair of horizontal mirror-image horseshoes, ⊂X⊃, with "X" representing *S*'s initial point of fixation. A scan to either the left or the right would result in exactly the same pattern of stimulation over time. This temporal equivalence is suggested by Deutsch as a possible explanation for the confusability of mirror-image stimulus pairs. In order for the two horseshoes to be distinguished spatially, additional information, provided by *S*'s knowledge of his direction of scan, would have to be integrated with this temporally organized stimulation. This analysis applies in the same way to mirror-image pairs in the vertical plane. For aligned pairs, however, no integration of motor scan knowledge is necessary since the directionality of these stimuli remains the same regardless of direction of scan.[4] If it is assumed that this integration of information from more than one source requires time, comparison of both horizontal and vertical mirror-image pairs should take longer than comparison of aligned pairs as found in the present study.

The difficulty young children have with mirror-image stimuli may be due in part to the requirement of integrating these sources of information. In addition, the fact that kindergarten children have unusual difficulty with horizontal mirror-image pairs compared with vertical, and with left–right stimuli compared with up–down (Huttenlocher, 1967a) suggests that at this age they cannot derive the necessary directional information from their movements in the horizontal plane. This lack of left–right response differentiation may well be, as Corballis and Beale (1970) have suggested, a function of the bilaterally symmetrical organization of the child's nervous system.

Gesell and Ames (1947) report that handedness does not become definitely established in the child until the age of five. The success of Caldwell and Hall's (1969) training procedures in decreasing confusion between mirror-image letter pairs is possibly explained by the fact that their kindergarten *S*s have just reached the age where directional differences, especially left–right discriminations, can be successfully processed when a training task relevant to these differences is used.

[4] The integration of visual with motor scan information has also been used by Ghent (1961) to explain children's choices of "right-side up" and "upside-down" stimuli, and by von Holst (1954) to account for the fact that the environment remains stationary during a voluntary eye movement.

REFERENCES

CALDWELL, E. C., & HALL, V. C. The influence of concept training on letter discrimination. *Child Development*, 1969, **40**, 63–71.

CALDWELL, E. C., & HALL, V. C. Distinctive-features versus prototype learning reexamined. *Journal of Experimental Psychology*, 1970, **83**, 7–12.

CORBALLIS, M. C., & BEALE, I. L. Bilateral symmetry and behavior. *Psychological Review*, 1970, **77**, 451–464.

DEUTSCH, J. A. A theory of shape recognition. *British Journal of Psychology*, 1955, **46**, 30–37.

GESELL, A., & AMES, L. B. The development of handedness. *Journal of Genetic Psychology*, 1947, **70**, 155–175.

GHENT, L. Form and its orientation: A child's eye view. *American Journal of Psychology*, 1961, **74** 177–190.

GIBSON, E. J., GIBSON, J. J., PICK, A. D., & OSSER, J. A developmental study of the discrimination of letter-like forms. *Journal of Comparative and Physiological Psychology*, 1962, **56**, 892–898.

HUTTENLOCHER, J. Children's ability to order and orient objects. *Child Development*, 1967, **38**, 1169–1176. (a)

HUTTENLOCHER, J. Discrimination of figure orientation: Effects of relative position. *Journal of Comparative and Physiological Psychology*, 1967, **63**, 359–361. (b)

MACKINTOSH, J., & SUTHERLAND, N. S. Visual discrimination by the goldfish: The orientation of rectangles. *Animal Behavior*, 1963, **11**, 135–141.

RIOPELLE, A. J., ITOIGAWA, J., RAHM, V., & DRAPER, W. A. Discrimination of mirror-image patterns by rhesus monkeys. *Perceptual & Motor Skills*, 1964, **19**, 383–389.

RUDEL, R. G., & TEUBER, H. Discrimination of direction of line in children. *Journal of Comparative and Physiological Psychology*, 1963, **56**, 892–898.

SEKULER, R. W., & HOULIHAN, K. Discrimination of mirror-images: Choice time analysis of human adult performance. *Quarterly Journal of Experimental Psychology*, 1968, **20**, 204–207.

SEKULER, R. W., & ROSENBLITH, J. F. Discrimination of direction of line and the effect of stimulus alignment. *Psychonomic Science*, 1964, **1**, 143–144.

SUTHERLAND, N. S. Visual discrimination of orientation and shape by octopus. *Nature*, 1957, **179**, 11–13.

VON HOLST, E. Relations between the central nervous system and the peripheral organs. *British Journal of Animal Behavior*, 1954, **2**, 89–94.

(Received May 13, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 273–279

PRIMARY TASK PERFORMANCE AS A FUNCTION OF ENCODING, RETENTION, AND RECALL IN A SECONDARY TASK [1]

DON TRUMBO [2] AND FRANCIS MILONE

Pennsylvania State University

Sixteen males were trained in pursuit tracking, then tracked while involved in encoding, retaining, or recalling sequences in a serial-learning secondary task. Results indicated that recall produced the greatest interference in the tracking task, followed by encoding, then retention. Tracking did not interfere with processing of the secondary task. In a second study, secondary task signals were presented aurally rather than visually with similar results. The findings are discussed in terms of attention demands and information-processing requirements underlying each phase of the secondary task.

In a recent article, Johnston, Greenberg, Fisher, and Martin (1970) reported four experiments in which a compensatory tracking task was combined in different experiments with verbal tasks whose primary information-processing requirements were encoding, retention, and recall, respectively. These dual-task studies indicated that tracking error was a function of the diffculty of the verbal task whether the task loaded on encoding, retention, or recall requirements. Furthermore, a rough indication of the divided attention demands was obtained by comparing the tracking error for the various experiments. These results suggested that recall of the second task engendered more interference with tracking than encoding, which in turn had greater effects than retention.

One difficulty with the conclusion drawn in the last statement was that the comparisons were made across different experiments using the same tracking task, but qualitatively different second tasks. Thus, as the authors acknowledged, these comparisons are tenuous and the conclusion needs further supportive evidence.

The evidence that encoding and retention tasks, as well as recall tasks, interfered with tracking also appears to contradict the results of a number of studies from our laboratory (Noble, Trumbo, & Fowler, 1967; Trumbo, Noble & Swink, 1967; Trumbo & Noble, 1970). In these studies, *S*s who were instructed to "attend to and learn" a series of secondary task signals (but with no overt response requirements) showed no interference in the overlapping tracking task. At the same time, evidence from test trials indicated that these *S*s did learn as much about the second set of signals as those *S*s who verbally anticipated each signal in the set.

It appeared that these "covert learning" conditions in our experiments were comparable to the encoding conditions in the Johnston et al. (1970) studies; yet while the latter tasks produced interference, the former did not. Whether this discrepancy resulted from the use of different tracking tasks (compensatory tracking of a complex sine wave in the Johnston et al. research vs. pursuit tracking of a step-function input in our studies), or the fact that Johnston et al. used a reward system biased in favor of performance on the verbal tasks, or from yet other differences in the second task requirements is not clear. One possibility is that the seemingly more difficult tracking task used by Johnston et al. failed to yield sufficient residual attention capacity for the encoding task, whereas with our simpler tracking task a sufficient residual remained to handle the less demanding encoding phase.

The present study was designed to assess the relative demands of successive stages

[1] This research was supported by Grant GB 6775 from the National Science Foundation. The authors are indebted to Merrill Noble for a number of helpful suggestions and to Fred Bates, Joe Propersi, and Randy Wilkinson, who assisted with data collection and analysis.

[2] Requests for reprints should be sent to Don A. Trumbo, Department of Psychology, Pennsylvania State University, 417 Psychology Building, University Park, Pennsylvania 16802.

of information processing in a secondary task, as indicated by performance decrements in an overlapping primary task, and, at the same time, to determine the effects of dual-task loading on the short-term memory for the secondary task. The paradigm differed from that of Johnston et al. and from our own earlier studies in that the tracking task was overlapped with different stages of processing on the same secondary task rather than on qualitatively different tasks.

Experiment I

Method

Subjects.—Sixteen male volunteers from introductory psychology classes received research participation credit for serving in the experiment.

Apparatus.—The tracking apparatus was very similar to that described in detail elsewhere (Trumbo, Eslinger, Noble, & Cross, 1963). Basically, it consisted of two *S* booths, each equipped with an oscilloscope on which target and cursor were displayed as $\frac{3}{4}$-in. vertical hairlines, and a steel chair with an arm control which consisted of a right forearm rest pivoted at the elbow and a grip handle, adjustable to the length of the forearm. The control was positional with a control-to-display ratio of 11° arc to 1-in. displacement. A stationary left armrest included four microswitches mounted to accommodate four fingers of the left hand. Four $\frac{1}{2}$-in. jeweled lamps positioned 4 in. apart above the cathode-ray tube (CRT) on an arc with a radius of 7 in. from the center of the CRT served as signal sources for the secondary task.

Signals for both the tracking task and the secondary tasks were programmed from a separate control room via a paper tape reader. Absolute integrated error was obtained for each tracking trial using an operational amplifier manifold and served as the criterion of tracking performance. Each of the four microswitches on *S*'s left armrest was wired to a separate pen of a multiple-pen event recorder from which the number of correct responses and the total response times were taken as indicants of secondary task performance.

Tasks.—The pursuit tracking task consisted of a step-function input with a sequence of eight discrete steps, 1 sec. apart, repeated six times per 48 sec. The secondary task consisted of a sequence of seven events programmed on the four jeweled lamps. These events occurred with an interstimulus interval (ISI) of 3 sec. and a duration of 1 sec/event. The *S*'s task was to encode, retain, and recall these seven events in successive 24-sec. intervals. The recall task was to reproduce the sequence using the four spatially corresponding buttons.

Design and procedures.—The design of the experiment is presented in Table 1. In Phase 1 of the study, all *S*s were given twenty 48-sec. tracking trials, separated by 18-sec. rest intervals and preceded by a 2-sec. warning buzzer. Integrated error scores were fed back to *S* via an intercom after each trial. At the completion of Phase 1, all *S*s were given a familiarization trial on the secondary task. They were presented with one sequence of seven lights which they were instructed to "attend to and learn" ("encoding"), followed by an unfilled retention interval of 24 sec. ("retention"); then on signal from *E*, *S*s attempted to reproduce the sequence by pressing the response keys ("recall").

Phase 2 consisted of the eight conditions shown in Table 1. Each row describes one trial in the design, with two *S*s beginning at each row and proceeding, in order, through the eight trials. Seven different sequences of the lights were generated, and all *S*s proceeded through the sequences in the same order. Thus, each sequence was used equally often with each of the first seven of the eight conditions, except that the two *S*s who began with the eighth (the tracking control) condition received the same combination of sequences and task conditions as the two *S*s who began with Cond. E_1. The eight trials (conditions) were separated by 1-min. intervals in which *E* described for *S* the manner in which the two tasks would be combined on the succeeding trial.

TABLE 1

Summary of Conditions, Experiment I

Cond.	Secondary task phase				
	Pre (24 sec.)	Encode (24 sec.)	Retain (24 sec.)	Recall (24 sec.)	Post (24 sec.)
E_1	—	Encode	Retain	Recall + track	Track only
E_2	—	Encode	Retain + track	Recall + track	—
E_3	—	Encode + track	Retain + track	Recall	—
E_4	Track only	Encode + track	Retain	Recall	—
E_5	—	Encode	Retain + track	Recall	—
E_6	—	Encode + track	Retain	Recall	—
CA	—	Encode (no tracking)	Retain (no tracking)	Recall (no tracking)	—
CB	—	Track only (48 sec.)	—	—	—

Note.—All *S*s pretrained with twenty 48-sec. tracking trials, then one secondary-task trial.

TABLE 2

MEAN INTEGRATED TRACKING ERROR, EXPERIMENT I

Tracking interval	Dual-task cond.						
	Track only	Track + retain	Track + encode	Track + recall	Track only (CB)	Track + retain	Track + encode
First 24 sec.	1.78 (E_4)	2.22 (E_2)	3.19 (E_3)	3.58 (E_1)	1.88 (CB)	2.19 (E_5)	3.16 (E_6)
Second 24 sec.	2.13 (E_1)	2.28 (E_3)	3.16 (E_4)	3.55 (E_2)	1.82 (CB)	—	—
$\bar{X}$	1.96	2.25	3.17	3.56	1.85	(2.19)	(3.16)

Note.—If S left the arm control alone and made no attempt to track, the error for a 24-sec. trial would be 6.31 v.

In Experimental Cond. E_1 through E_4 and Control Cond. B, each tracking trial was divided into two 24-sec. periods with a 5-sec. interval between periods to allow E to record the error for each period. Experimental Cond. E_5 and E_6 with only one 24-sec. tracking period were included in order to provide unequivocal evaluation of the effects of tracking on the retention (E_5) and encoding (E_6) phases as reflected in secondary task recall scores.

The instructions for Phase 1 described the requirements of the tracking task, illustrated typical errors (overshooting, undershooting, lagging, excessive leading, and slow rate of movement) using prototype tracking records, and emphasized the need for anticipation and rapid movements to minimize tracking error. Following Phase 1, Ss were instructed in the secondary task. They were told to attend to and learn the sequence of lights which would be presented, and, on signal from E, to reproduce the sequence on the response keys. Preceding each of the eight test trials, the sequence of events for the following trial was described and illustrated by means of a simple diagram of the appropriate row from Table 1.

Results

Primary task performance.—Of principal interest with respect to performance on the tracking task were Experimental Cond. E_1 to E_4. In these conditions, tracking was performed alone and in combination with each phase of processing of the secondary task, once in the first 24 sec. and once in the second 24 sec. of tracking. Therefore, a $2 \times 4 \times 8$ analysis of variance was performed on these cells, with within-Ss main effects of (*a*) first versus second tracking intervals and (*b*) the four dual-task conditions, and between-Ss main effect for the eight pairs of Ss who received the treatments in the same orders.

Neither the variance between Ss nor any of the interactions involving Ss was significant. Furthermore, the variance between first and second tracking trials failed to approach significance ($F < 1.00$). However, the variance attributable to dual-task conditions was significant, F (3, 24) = 65.29, $p < .01$. The mean integrated error scores for these conditions, as well as Control Cond. B and Experimental Cond. E_5 and E_6 are presented in Table 2. In addition to the differences attributable to dual-task conditions, perhaps the most striking fact about these data is the high degree of consistency within the columns, that is, across replications of the same combination of primary and secondary task conditions. The single trial conditions, E_5 and E_6, may be compared with Columns 2 and 3, respectively, and again the indications are that the results are highly consistent.

The means of Columns 1–4 of Table 2 were tested for separation using the Duncan's new multiple-range test. The results indicated that all the comparisons among the four means were reliable ($p < .05$).

The results are presented graphically in Fig. 1. The bars represent the mean integrated error scores from Table 2, rearranged to emphasize the amount of replication within the study. With the exception of the E_1 data for tracking only, estimates of the same dual-task effects are within .1 volt of one another.

Secondary task performance.—Two indicants of secondary task performance were analyzed: the number of errors in recalling the sequence of seven events and the total time to recall, measured from the first to the last response. Since Ss were not instructed to produce recall responses as quickly as possible, the latency scores were

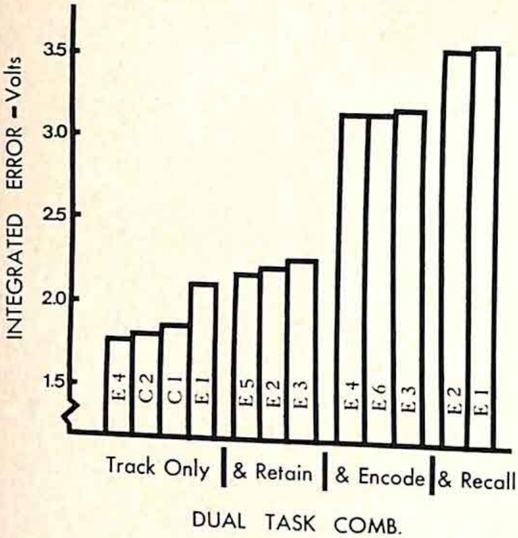

FIG. 1. Tracking error as a function of dual-task conditions. (Bars are coded with reference to conditions shown in Table 1.)

not considered critical measures of task performance.

The sequences of button presses recorded on the Esterline-Angus were scored independently by three judges who then reached a consensus score for each trial. Since recall was the "free ordered" type, that is, ordered recall but self-paced and without serial cues, scoring complexities occurred. Thus, for example, *S* might reproduce the first three signals correctly, omit the fourth, and then reproduce the last three signals correctly. A conservative scoring procedure would yield a score of three in this instance, whereas a more qualitative analysis would credit *S* with six correct responses. The procedure used was to score *S*'s sequence of response from both ends, crediting him with all responses corresponding to the signal sequence from the beginning, and then adding the number of responses corresponding to signals reading backward from the end of the sequence. Total response times were obtained by measuring the distance between the onset of the first and last responses from the response records.

The mean number of correct responses for recall of the seven-item secondary task sequences varied from 5.06 for Cond. E_3 (encode and retain while tracking) to 5.88 for Control B (secondary task only). The three conditions which involved encoding while tracking (E_3, E_4, and E_6) yielded the poorest performance, followed by the two conditions requiring recall while tracking (E_1 and E_2), retention while tracking (E_5), and the control. However, while these data show an interesting orderliness, an analysis of variance indicated that the array of means did not differ reliably, F (6, 90) < 1.00. Similarly, while the total response times were greatest for the two conditions in which responses were made during tracking (E_1 and E_2), the between-treatments variance was not significant, F (6, 90) $= 1.28$, $p > .05$. Thus, the data failed to establish any reliable effects of the primary task on secondary task performance.

Discussion

The results for the tracking task are qualitatively similar to those by Johnston et al. (1970); the greatest amount of interference occurred under dual-task conditions involving the recall phase of the secondary task followed by the encoding and retention phases. The fact that the latter two phases were associated with reliable decrements in tracking performance is also in agreement with the Johnston et al. findings, but appears contrary to our own prior results in which stimulus encoding tasks did not produce interference, regardless of whether the primary task was tracking (Noble et al., 1967; Trumbo et al., 1967) or serial verbal learning (Trumbo & Noble, 1970). If our assumption that these various secondary task conditions are comparable with respect to basic information-processing requirements is defensible, then the problem is to account for the discrepancy between present findings and those of Johnston et al. on the one hand and the results of our prior studies on the other hand.

A possible explanation may be found in the mode of presentation of secondary task signals. In our prior studies, secondary task signals were presented aurally, while primary task input was visual. The only exception to this was the case in which the primary task was serial verbal learning (Trumbo & Noble, 1970). In the latter, lists were presented on a memory drum with an ISI of 3 sec., and secondary task signals were presented via a set of five lights set in an arc above the exposure window. This arrangement was very similar to that in the

present study, with the exposure window replaced by the CRT display. In either case, the secondary task signals were readily discriminable in peripheral vision, i.e., without shifting focus from the primary task display. However, it is conceivable that in both studies *S*s did shift their point of focus from primary to secondary displays in response to the onset of the secondary task signals. While a momentary shift of the eyes might well have little or no effect on the discrete verbal learning task, it is quite possible that such a disruption of eye–hand coordination in the continuous adjustment tracking task would be detrimental to tracking performance. Thus, the discrepancy between the present results and those from prior studies may have been the result of the eye movements occasioned by the tendency to orient toward the source of secondary task signals.

Experiment II was designed primarily to test the above explanation of the results of Exp. I. In addition, since the prior experiments had required in some cases verbal and in other cases motor responses, these two response modes were included in the design. It also was decided to use an independent groups design for comparison with the within-*S*s design of Exp. I, and, finally, to run repeated trials under each experimental condition to examine the persistence of any observed performance decrements.

Experiment II

Method

Subjects.—Fifty-three male volunteers from introductory psychology classes received research participation credit for serving as *S*s. Data for 3 of these *S*s were discarded because of *E* errors or apparatus failures.

Apparatus.—The apparatus and the input for the tracking task were identical with those used in Exp. I. For the secondary task, a nine-item sequence of the numbers 1–4, generated with the same restrictions as the sequences in Exp. I, replaced the sequences of seven light signals used in Exp. I. These numbers were recorded on an audiotape system with an ISI of 2.0 sec. Each *S* booth was equipped with four response buttons, as in Exp. I, a remote intercom station for communication with *E*, and a headset for receiving secondary task signals.

Design and procedures.—The design was a 2 × 3 factorial with 2 response modes (motor vs. verbal) and 3 dual-task phasing conditions (encoding, retention, and recall). Five *S*s were assigned unsystematically to each cell, except that *S*s scheduled in pairs were assigned to the same dual-task condition, with one assigned to the motor and the other

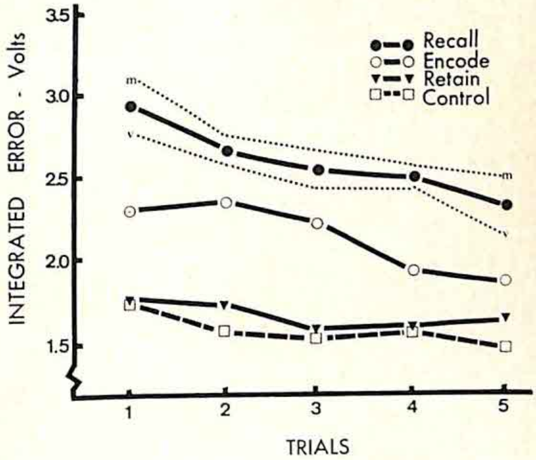

FIG. 2. Tracking error for dual-task and control conditions at each trial, Exp. II. (Data are pooled for verbal and motor response modes. Dotted lines show verbal, *V*, and motor, *M*, data separately for recall only. Tracking error would have been 6.31 v. if *S* had made no attempt to track.)

to the verbal response mode. Ten *S*s were assigned to a tracking control condition, and 5 each to two secondary task controls (motor and verbal).

Following initial tracking instructions, all *S*s received 30, 24-sec. tracking trials separated by 12-sec. rest intervals and anticipated by a 2-sec. warning buzzer. Integrated absolute error scores were reported to *S* after each trial via the intercom. Upon completion of the 30 tracking trials, *S* was given instructions and a single familiarization trial on a secondary task sequence without tracking. Then the phasing of the two tasks was explained and *S* was given five dual-task trials, wherein the tracking task was the same as that used in the initial training but the secondary task sequence differed from that used in the familiarization trial.

The total cycle for the secondary task was 90 sec., with 16 sec. for presentation of the numbers, a 44-sec. retention interval followed by the verbal command "recall," 24 sec. (maximum) for recall, and 6 sec. between the end of the recall interval and the next presentation. For the three dual-task conditions, respectively, the 24-sec. tracking trials overlapped the encoding, retention, or recall phase of the secondary task. Thus, within the five-trial block, tracking trials were separated by 66 sec.

Results

The data for the primary (tracking) task are summarized in Fig. 2. Mean integrated absolute error scores are plotted for Trials 1–5 as a function of dual-task conditions. It is apparent from these data that interference in tracking performance was

greatest when combined with the recall phase of the second task, next greatest for the encoding phase, and minimal for the retention phase of the second task. The motor recall mode appeared to produce more interference than the verbal, but only when the recall phase was combined with tracking. That is, the effects of encoding or retaining the sequence of numbers appeared not to depend on the response code required.

These findings were supported by a $4 \times 2 \times 5$ analysis of variance involving the 3 dual-task conditions plus the tracking control and the 2 response modes as between-*S*s sources of variance and the 5 trials as repeated measures. The dual-task combinations were a reliable source of variance, $F(3, 32) = 7.85$, $p < .01$, but the difference between response modes was not significant, $F < 1.00$. Trials was a significant source of variance, $F(4, 128) = 10.63$, $p < .01$, but neither the interaction of trials and dual-task combinations ($p > .10$) nor any of the remaining interactions was significant.

A Duncan's new multiple-range test indicated that the tracking-plus-recall and tracking-plus-encoding conditions differed reliably ($p < .05$) from both the tracking control and tracking-plus-retention conditions, but that the remaining differences were not reliable.

Recall scores (number of correct responses) for the secondary task again suggested that tracking interfered most with processing of the number sequences when it overlapped the encoding phase of the task. However, two analyses of variance failed to provide statistical support for this trend. The first analysis was the same as that for tracking error ($4 \times 2 \times 5$), but only the main within-*S*s effect of trials was significant. Since the group means tended to converge over trials, a second analysis (4×2) was run on the data for Trial 1, where the variance due to dual-task conditions did not reach significance ($p = .12$).

Discussion

The results of Exp. II provide no support for our speculation that the interference from encoding in the first experiment might have been an artifact of eye movements between primary and secondary task signal sources. Instead, they indicate that both encoding and recall phases of processing the second task are sufficiently demanding of attention that they interfere with the tracking task, regardless of whether the signals are visual or auditory. In this sense, the two experiments are complementary in increasing the generality of the findings. Generality is further increased in that the present results are in substantial agreement with those of Johnston et al. (1970), which involved rather different tasks and procedures. The findings are somewhat less consistent with respect to the attention demands of retention, although it appears that processes occurring during a retention interval are less demanding than either encoding or recall.

The results of the first experiment did not indicate any accumulative effects on tracking performance of overlapping with two phases of the second task. That is, tracking scores during retention of the second task were nearly identical whether they were preceded by encoding only or encoding plus tracking. Similarly, tracking scores during recall appeared to be independent of whether or not the prior retention phase was combined with tracking. These findings appear to be consistent with the evidence that tracking did not interfere appreciably with performance on the secondary task.

Neither experiment yielded reliable evidence that tracking interfered with performance on the secondary task or was more detrimental to one phase of secondary task processing than to another. Furthermore, there was no evidence of cumulative effects when tracking overlapped two, as compared with one, phases of the second task. However, in both studies there was some indication that tracking might affect the encoding process adversely. A more sensitive measure of secondary task performance, less susceptible to chance variability, might substantiate this trend. Nevertheless, the present results indicate no differential interference with the phases of the secondary task, and therefore support the contention that the attention demands of the task phases are reflected in tracking task performance.

Taken together with the findings of Johnston et al. (1970), the results indicate that response selection and/or response execution, which are obviously required at recall, are either more demanding stages of information processing or more limited in capacity than the stimulus preprocessing or stimulus categorization required

at encoding. In this respect, the results agree with conclusions reached on the basis of our prior studies (Noble et al., 1967; Trumbo & Noble, 1970; Trumbo et al., 1967) to the effect that response-selection decisions constitute a major limitation to the processing of overlapping channels of information. However, the present results do not agree with those prior studies which indicated that encoding tasks did not exceed reserve attention capacity and therefore did not interfere with the tracking. Further research directed at this discrepancy may be worthwhile.

Generalizations about the attention demand of encoding, retention, and recall, or the information-processing stages underlying these task phases, should be most tentatively drawn in lieu of more adequate sampling of both primary and secondary task conditions. It may be, for example, that recall produced more interference in the present studies simply because *both* primary and secondary tasks were relatively demanding at the response-selection and/or response-execution stages of processing. The experimental paradigms used in the present studies appear to offer a promising strategy for pursuing these and related questions.

REFERENCES

Johnston, W. A., Greenburg, S. N., Fisher, R. P., & Martin, D. W. Divided attention: A vehicle for monitoring memory processes. *Journal of Experimental Psychology*, 1970, **83**, 164–171.

Noble, M., Trumbo, D., & Fowler, F. Further evidence of secondary task interference in tracking. *Journal of Experimental Psychology*, 1967, **73**, 146–149.

Trumbo, D., Eslinger, R., Noble, M., & Cross, K. A versatile electronic tracking apparatus (VETA). *Perceptual & Motor Skills*, 1963, **16**, 649–657.

Trumbo, D., Noble, M., & Swink, J. Secondary task interference in the performance of tracking tasks. *Journal of Experimental Psychology*, 1967, **73**, 232–240.

Trumbo, D., & Noble, M. Secondary task effects on serial verbal learning. *Journal of Experimental Psychology*, 1970, **85**, 418–424.

(Received May 26, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 280–286

S–R COMPATIBILITY AND THE RELATIVE FREQUENCY EFFECT IN CHOICE REACTION TIME

HAROLD L. HAWKINS[1] AND JAMES R. UNDERHILL

University of South Florida

The effects of relative stimulus frequency on choice reaction time (RT) were investigated under two levels of S-R compatibility. Choice RT increased with declining frequency and the magnitude of this increase remained invariant across levels of compatibility when relative frequency was manipulated by varying the frequency imbalance among stimulus alternatives. However, when relative frequency was manipulated by varying the number of equally likely alternatives, the increase in choice RT associated with declining relative frequency was significantly greater under conditions of low, relative to high, compatibility. These findings suggest that the relative frequency effect is produced by two distinctly different mechanisms of anticipatory bias. One of these functions at stimulus identification and is responsive to stimulus relative frequency, whereas the other functions at response selection and is primarily sensitive to response numerosity.

It has been well-established that choice reaction time (RT) varies as the inverse of the relative frequency of individual alternatives when stimuli and responses are paired in 1:1 correspondence (e.g., Fitts, Peterson, & Wolpe, 1963; Leont'ev & Krenchik, 1963). Until recently, however, it was unclear whether this relationship, the so-called relative frequency effect, is produced by variations in stimulus probability or by variations in response probability since these two variables are completely confounded when stimuli and responses are paired in the 1:1 manner. One approach to the solution of this difficulty has been the use of many:1 S-R correspondences whereby stimulus and response probability can be independently manipulated. The results of studies taking this approach have tended to support the conclusion that functional stimulus probability is the primary determinant of the relative frequency effect (Bertelson & Tisseyre, 1966; Hawkins & Hosking, 1969; Hawkins, Thomas, & Drury, 1970).

A theoretical issue underlying much of the research on the relative frequency effect concerns the significance of two logically distinct decision-making activities commonly supposed to intervene between the stimulus and the response in choice RT. Upon stimulus input, *S* presumably decides which of the several possible stimuli has been presented (stimulus identification, or discrimination) and then which of the several permissible response alternatives is appropriate for that stimulus (response selection). On the assumptions that stimulus probability effects are a reflection of bias in stimulus identification and that response probability effects are a reflection of bias in response selection, the foregoing evidence is consistent with the view that the relative frequency effect is largely due to anticipatory adjustments in the stimulus identification process. The present study constitutes a further test of this view.

If the slope of the function relating choice RT and the relative frequency of alternatives, in fact, is governed wholly by processes involved in stimulus identification, then that slope should be exclusively responsive to manipulations of the stimulus array and should not be responsive to variations in the characteristics of the response display per se or, more particularly, to variations in the compatibility (degree of familiarity) of the S-R pairings under study. In a review of the human performance literature, Posner (1966) has summarized findings seemingly at variance with this prediction. Posner's evidence was based on a comparison of the results of a number of choice RT experiments containing S-R pairings distributed across a

[1] Requests for reprints should be sent to Harold L. Hawkins, Department of Psychology, University of South Florida, Tampa, Florida 33620.

broad range of compatibility levels. The comparison revealed that as compatibility declines there is a concomitant acceleration in the rate at which RT increases with the number of equally likely stimulus alternatives (ELAs). Since the relative frequency of individual alternatives necessarily declines with increases in the number of ELA, Posner's analysis suggests that the relative frequency effect is not independent of S-R compatibility. Clearly, this finding makes it difficult to defend the general proposition that the relative frequency effect can always be attributed to anticipatory bias in stimulus identification, regardless of the specific procedure employed to manipulate relative frequency. However, these findings do not preclude the possibility that stimulus identification plays the significant role in the relative frequency effect as induced by alternative means, viz., manipulation of the frequency imbalance among a set number of unequally likely stimulus alternatives (ULAs). While it is true that ELA and ULA procedures yield similar results under certain circumstances, e.g., when they serve as alternate means for varying the total amount of information in a stimulus series, they would not appear to be influenced in the same way by variations in S-R compatibility. Hyman (1953), for instance, has reported data which may be interpreted as showing that under conditions of low compatibility, ELA procedures produce a stronger relative frequency effect than do ULA procedures. However, the reverse is true under conditions of high compatibility, as indicated by a comparison of results between two studies of naming RT, one by Brainard, Irby, Fitts, and Alluisi (1962) and the other by Fitts et al. (1963). Although Brainard et al. found that stimulus relative frequency has only modest effects on naming RT when the number of equiprobable numeral stimuli is varied, Fitts et al. found a sizable difference in naming RT to the same class of stimuli under conditions of frequency imbalance. Considered together, these findings suggest that ULA procedures either are insensitive to variations in S-R compatibility or, perhaps, that they are less sensitive than are ELA procedures. Only the first of these two possibilities is totally consistent with even a limited conception of stimulus identification as a necessary factor in the determination of the relative frequency effect.

With these considerations in mind, the present study was designed to vary relative frequency both by manipulating the number of ELAs and by manipulating the frequency imbalance among ULAs. As a test of the stimulus identification interpretation of the relative frequency effect, two conditions of S-R compatibility were studied. Under the high compatibility condition, the alphabetic symbols used as stimuli were assigned naming responses, and in the low compatibility condition these same stimuli were assigned key-press responses. Under all conditions of the experiment, stimuli and responses were paired on 1:1 basis.

Method

Subjects.—The *Ss* were 7 women and 5 men enrolled in an advanced undergraduate experimental psychology course. All *Ss* had normal or normal corrected vision and all were naïve with regard to RT research. Course credit was given for participation.

Stimuli and apparatus.—The stimulus population consisted of the uppercase letters A, D, E, H, Q, S, X, and Z (Tactype No. 5418). Stimuli were presented singly for a duration of 1 sec. on a uniform white background at a viewing distance of 43.18 cm. through a Polymetric two-field mirror tachistoscope. Response latencies were recorded with a Hunter Klockounter. Two separate response modes were distinguished. Under conditions calling for the key-press response, a response panel containing eight piano-type keys was placed before *S*. The 4 keys assigned to each hand were arranged in an inverted U configuration so that the thumb and first three fingers could be comfortably rested on them. A slight depression of a key closed a relay contact which, in turn, (*a*) terminated the Klockounter and (*b*) activated one of a 2×4 array of indicator lights informing *E* of the identity of the particular key depressed. Under conditions requiring the naming response, a microphone was suspended about 2.54 cm. in front of *S* at mouth level. Vocalizations into the microphone closed a Marietta voiced-activated relay which in turn terminated the Klockounter.

Procedure.—The 12 *Ss* were randomly divided into two groups of six. Throughout the one practice and six 1.25-hr. test sessions forming the experiment, one of these groups was assigned the key-press response mode and the other was assigned the naming

response mode. The practice and first three test sessions were presented on consecutive days and the last three sessions followed on consecutive days after a 2-day delay. Three days prior to the practice session, *S*s were given material which they were instructed to memorize prior to the beginning of the experiment (which they did). Each member of the key-press group was given a card on which were depicted the eight S-R pairings specifically assigned to that *S* and the spatial arrangement of the response keys. Each member of the naming group was given a list containing the eight stimulus letters to be used in the experiment.

During the 240-trial practice session, *S*s in both groups were trained under a general procedure in which any one of the eight stimuli could appear with equal probability on any trial. The *S*s were instructed here, as on subsequent days, to respond as quickly and accurately as possible. The specific S-R pairings assigned to *S* during practice were maintained across days of the experiment. In addition, all *S*s practiced the response mode they would be using on subsequent days.

During each of the six experimental sessions, *S* was tested for 40 trials on each of six conditions of relative stimulus frequency. Three of these conditions differed with respect to the number of equally likely alternatives and three differed with respect to the frequency imbalance among alternatives. The three frequency imbalance conditions were designed as, I: 25(A)/75(B); II: 87.5(A)/12.5(B); and III: 25(A)/50(B)/12.5(C)/12.5(D). The Roman numeral beginning each sequence is an arbitrary designation which will be used hereafter in referencing each of the various conditions. Each digit code within a sequence indicates the percentage of trials on which a particular S-R pairing occurred. The letter following each percentage value in each sequence arbitrarily designates the identity of the S-R pairing which took on that relative frequency of occurrence. Thus, e.g., for a particular *S*, the letter A (in the above sequences) might represent the stimulus "X" (requiring, e.g., the left index finger response under the low compatibility condition), which occurred with probabilities .25, .875, and .25 across the three conditions of frequency imbalance. In terms of the above definitions, the three equiprobable stimulus conditions were: IV: 50(A)/50(B); V: 25(A)/25(B)/25(C)/25(D); and VI: 12.5 (A)/12.5 (B)/12.5 (C)/12.5 (D)/12.5 (E)/12.5(F)/12.5(G)/12.5(H).

Within each response mode, the order of presentation of the six conditions was completely counterbalanced within *S*s across days and across *S*s within days. The assignment of the stimuli designated as A and B to index fingers of the dominant and nondominant hands was counterbalanced across *S*s within the key-press condition. The stimuli denoted as C and D were assigned to the middle fingers, E and F to the thumbs, and G and H to the right fingers. The identity of the stimulus designated as A, etc., was determined on a separate random basis for each *S*.

During both practice and test sessions, the identity and correct relative frequencies of alternatives and, in the case of the key-press group, the appropriate S-R code, were made available to *S* on an index card prior to the onset of each condition. The card remained before *S* throughout a condition. The intertrial interval was 5 sec., with a warning tone of 1-sec. duration occurring immediately before each presentation of a stimulus. After each correct response, *S* was informed of his RT in milliseconds. Following each incorrect response, *S* was told only that he had comitted an error. Stimuli to which *S* incorrectly responded were *not* subsequently repeated within a condition.

Results

Table 1 gives the mean latencies and percent incorrect responses under each condition. Except in the case of Cond. ULA III, the RTs appearing in Table 1 represent the pooled arithmetic mean latencies for S-R pairs denoted as A and B. The pooling of times for A and B, a procedure designed to increase the reliability of our estimates, was justified by the fact that the two sets of times did not differ reliably, $t\ (11) = .56$, $p > .10$. To assess the effect within Cond. ULA III of including data from the less frequent C and D, the mean latencies of these S-R pairs pooled across Cond. ELA V and ELA VI were compared with those for Pairs A and B. Under conditions of low compatibility, the pooled RTs for A and B was 597 msec. and for C and D was 608 msec., $t\ (5) = .47$, $p > .10$. Under the high compatibility condition, the two values, respectively, were 443 and 450 msec., $t\ (5) = .31$, $p > .10$. Because of the lack of significance between these values, the assumption was made that the latencies reported under ULA III for C and D could be fairly compared with those for A and B under that or any other condition.

In what immediately follows, the three ELA conditions were treated within a single analysis and the three ULA conditions were each treated separately. Separate analyses were necessitated by the fact that levels of relative frequency were not always directly comparable across conditions. All analyses followed the mixed model, with S-R compatibility (high vs. low) as a between-*S* variable and relative frequency and level of practice (Days 1

TABLE 1

CHOICE RT (IN MSEC.) AND ERROR PERCENTAGE AS A FUNCTION OF STIMULUS RELATIVE FREQUENCY, LEVEL OF S-R COMPATIBILITY, AND METHOD OF FREQUENCY MANIPULATION

Cond.	High compatibility					Low compatibility				
	Stimulus relative frequency					Stimulus relative frequency				
	.875	.75	.50	.25	.125	.875	.75	.50	.25	.125
ULA I	—	378 (.00)	—	428 (.01)	—	—	384 (.70)	—	436 (2.33)	—
II	396 (.00)	—	—	—	439 (.00)	394 (.48)	—	—	—	434 (3.46)
III	—	—	394 (.00)	442 (.01)	459 (.06)	—	—	503 (1.67)	542 (2.99)	575 (4.90)
ELA IV, V, VI	—	—	402 (.00)	432 (.00)	458 (.09)	—	—	420 (.72)	540 (1.80)	654 (3.10)

and 2 vs. 3 and 4 vs. 5 and 6) as within-*S* variables. The between-*S* variable was not of direct interest as a main effect, it should be pointed out, because of differences across compatibility levels in the apparatus mediating between *S*s' response and Klockounter termination. The effect of this difference in equipment was to add (or subtract) a constant of unknown magnitude to all times obtained under the high, relative to low, compatibility task. This factor invalidated main effect comparisons on the compatibility variable, but not evaluations of interactions involving this variable.

The analysis of ELA conditions revealed that RT declined with increasing compatibility, $F\ (1, 10) = 35.52$, $p < .005$, with increasing relative frequency, $F\ (2, 20) = 96.96$, $p < .005$, more rapidly with increasing relative frequency under the low relative to high compatibility condition, $F\ (2, 20) = 39.30$, $p < .005$, and with increasing practice, $F\ (2, 20) = 7.99$, $p < .005$. The slope of the fitted regression line across high compatibility ELA conditions was 1.52 msec., and across low compatibility ELA conditions was 6.40 msec., per unit change in stimulus probability. The significant interaction of relative frequency and compatibility obtained in the analysis of ELA conditions indicated that the differences in slope across compatibility levels were statistically reliable.

Under ULA I, RT declined with increasing relative frequency, $F\ (1, 10) = 75.60$, $p < .005$, and with increasing level of practice, $F\ (2, 20) = 12.51$, $p < .005$. Under ULA II, RT again declined with increasing relative frequency, $F\ (1, 10) = 35.60$, $p < .005$, and increasing level of practice, $F\ (2, 20) = 5.96$, $p < .01$. Under the last condition, ULA III, RT declined with increasing compatibility, $F\ (1, 10) = 20.99$, $p < .005$, increasing relative frequency, $F\ (2, 20) = 31.63$, $p < .005$, and increasing level of practice, $F\ (2, 20) = 7.30$, $p < .005$. A significant level of Practice × Relative Frequency interaction, $F\ (4, 40) = 12.08$, $p < .005$, also was obtained under ULA III, indicating a decline in the magnitude of the relative frequency effect across levels of practice. The slopes of the fitted regression lines for the high and low compatibility conditions, respectively, were 1.00 and 1.04 under ULA I, .92 and .70 under ULA II, and 1.79 and 1.86 under ULA III. The analysis of variance carried out on data from ULA conditions revealed that none of these differences in slope within ULA conditions across levels of compatibility were statistically reliable.

A separate analysis was carried out to compare performance across high compati-

bility ELA Cond. IV, V, and VI with that obtained under the ULA condition, (III), containing equivalent relative frequency values. In this analysis, method of frequency manipulation (ELA vs. ULA), relative stimulus frequency (12.5 vs. 25 vs. 50), and level of practice were all treated as within-*S* variables. The significant effects of relative frequency and of level of practice under Cond. III, IV, V, and VI have been reported immediately above and will not be described again here. Our major concern in the present analysis centered around the main effect of method of frequency manipulation and the interaction of this variable with relative frequency. The *F* ratios for both these effects were nonsignificant.

Discussion

The results of this experiment are quite clear. When stimulus frequency is manipulated by varying the frequency imbalance of alternatives, the rate of increase in choice RT with decreasing relative frequency is unaltered by variations in S-R compatibility. However, when relative frequency is varied by increasing the number of equally likely alternatives, the rate of increase RT is significantly greater under conditions of low relative to high compatibility.

Two basic assumptions underlie our analysis of these results. First, we have assumed that the interval between stimulus and response in choice RT is partially taken up by two logically distinct processes, one relating to stimulus identification and the other to response selection. Second, we have assumed that the locus, relative to the aforementioned processes, of the anticipatory bias underlying the relative frequency effect can be assessed on the basis of whether or not the effect is influenced by S-R compatibility. Considered in these terms, the results of this study suggest that the effect of frequency imbalance on RT results exclusively from anticipatory adjustments in the stimulus identification phase of information processing. To this extent, the present findings are in essential agreement with earlier data from this laboratory concerning the effects of frequency imbalance on RT (Hawkins & Hosking, 1969; Hawkins, Thomas, & Drury, 1970). However, a different picture emerges on considering performance under conditions where the number of ELA were varied. The interaction of compatibility and relative frequency obtained here indicates that under certain conditions, factors related to response selection play an important role in determining the relative frequency effect. Put differently, the rise in choice RT observed with increases in the number of equiprobable stimulus alternatives would not appear to be due solely to concomitant changes in the stimulus processing demands of the task because the magnitude of the rise was largely dependent upon the compatibility of the S-R code used. Indeed, the change which took place in the slope of the ELA relative frequency function across levels of compatibility occurred in spite of the fact that the identity and statistical structure of the stimulus array remained invariant across these levels.

Considered altogether, the findings of this experiment suggest that the relative frequency effect may be the result of two distinctly different information-processing mechanisms, one functioning during stimulus identification and the other during response selection. A number of alternative theoretical approaches are available in terms of which these two decision mechanisms can be conceptualized (see, e.g., Berlyne, 1957; Smith, 1968). However, we have elected to interpret our findings in terms of an informal model derived from statistical decision theory, primarily because this approach offers a single framework within which both the stimulus and the response effects obtained in the present experiment can be adequately described. The stimulus identification component of the model is based in part on a view first described by Stone (1960). The description of the response selection component borrows from the formulations of Keele (1966) and of Berlyne (1957). The central feature of the model of stimulus identification is an internal array of counters or data accumulators each of which is maximally responsive to a particular task-relevant stimulus alternative among those potentially available in the external environment. While each counter is primarily responsive to a particular stimulus alternative, counters are also generally responsive to randomly varying levels of neural noise assumed to exist as an inherent property of the human sensory system. The effect of presenting a given stimulus is thus to increase the rate of accumulation of counts in the appropriate counter beyond the average rate of accumulation of noise-produced counts in the remaining counters. When the count in a particular counter (whether correct or in-

correct) reaches a predetermined criterion level, the identification response associated with that counter is initiated. In decision models of this general form, the stringency of the criterion setting is a major determinant of the expected length of time needed to sequentially sample data counts from the external stimulus. The more stringent the criterion, the higher the count necessary to reach that criterion, and thence, the longer the data sampling time required. Conversely, the more lenient the criterion, the shorter the sampling time. However, the likelihood that the counter, when incorrect, will inadvertently reach criterion on the basis of noise alone also becomes greater as the criterion becomes more lenient. Thus, the locus of the criterion will depend in part on the speed-accuracy emphasis imposed by the conditions under which the decision-maker is performing. An identification mechanism of this type can be structured so that it adaptively reacts to differences in stimulus relative frequency by establishing differentially stringent criteria across counters. More particularly, the decision mechanism should operate in such a way that the stringency of the criterion established for a given counter will vary as the inverse of the relative frequency of the associated stimulus alternative. One implication of this assumption is that the effects of varying stimulus relative frequency on the speed of stimulus identification should be independent of whether relative frequency is varied by ULA or by ELA procedures.

On this basis, the fact that the magnitude of the relative frequency effect, as measured across the high compatibility ELA Conds. IV, V, and VI, approximated that obtained under ULA Cond. III (which involved relative frequency values equivalent to those occurring across the ELA conditions) indicates that the effect in the former instance may be wholly attributable to anticipatory bias in stimulus identification. As previously indicated, however, the unique feature of the ELA data is the interaction of compatibility and relative frequency about which models focusing on the stimulus identification phase of information processing have little to say. We view this interaction as a manifestation of response conflict (Berlyne, 1957) which, we assume, increases as a monotonic function of the number of response alternatives available within the RT task. The extent of this increase in conflict, we further assume, is dependent upon the strength of association between stimuli and their respective responses. More specifically, it is proposed that response conflict increases with the number and extent of competing response tendencies elicited by each stimulus. To the extent that a set of S-R pairings is strongly associated, as under our high compatibility naming condition, only the correct response will tend to be elicited by each stimulus. Accordingly, such pairings will normally produce relatively little conflict. However, when S-R pairings have been associated only marginally, as in our low compatibility key-press condition, each stimulus alternative will elicit response conflict by evoking responses other than that experimentally defined as appropriate. To summarize, the high and low compatibility tasks studied in this experiment seem to be equally sensitive to the effects of changing stimulus relative frequency, regardless of how such changes are experimentally achieved. Furthermore, performance on the low compatibility task appears to be additionally sensitive to a form of response conflict that is associated with variations in the number of responses *S* is required to discriminate.

While response conflict seems to be a reasonable label for the response numerosity effects most evident under the low compatibility task, the concept does not imply a concrete mechanism by which such effects occur. Statistical decision theory provides a conceptual system within which a relatively specific mechanism of conflict can be formulated. According to one version of such a mechanism, associated with each possible response alternative is a counter which receives and accumulates more or less noisy data representing the output of the stimulus identification process described earlier. When the count in one of these response counters, whether correct or incorrect, reaches a predetermined criterion, the associated motor response is activated. The effect of S-R learning, or compatibility, can be represented within this scheme in terms of the average difference in rate of count accumulation between correct and incorrect counters. That is, increasing compatibility corresponds to increased signal strength, or in present terms, to increases in the rate at which the count accumulates in the correct counter. Thus, at relatively high levels S-R learning, as in our high compatibility task, the rate at which the signal plus noise-produced count builds in the appropriate response counter will tend to be much greater than the rate of accumulation of noise-produced counts in the incorrect counters. At lower levels of S-R

learning, as in our low compatibility task, signal strength is less and therefore the difference between correct and incorrect counters in the rate at which the count builds will tend to be less. Following the usual statistical decision theoretical assumption that noise (or the rate of accumulation of noise-produced counts) is distributed across counters as an independent random variable, the smaller the average difference in rate of count accumulation between correct and incorrect counters, the greater the likelihood at any criterion setting that a member of the latter set will inadvertently reach criterion prior to the former. Moreover, increases in the number of incorrect response counters (as across ELA conditions) should, for any given criterion setting, lead to an increase in the likelihood that one of the independently accumulating incorrect counters will perchance reach criterion first. The extent of the increase in error rate will depend on S-R compatibility. The greater the difference between correct and incorrect counters in the rate at which they accumulate counts, the less the effect on error rate of an increase in the number of incorrect response alternatives. Actually, however, we normally do not permit our *S*s to react to increases in the number of S-R alternatives by relaxing accuracy while maintaining response latency at a constant level. Instead, we insist that error rate remain within reasonable limits so that we can directly observe the effects of our task manipulations on response latency. From the present point of view, this requirement is met by means of an increase in criterion stringency, and thence, by an increase in the time necessary to sequentially sample data from the stimulus identification mechanism. It is this increase in sampling time which, we have shown, is dependent upon the strengths of the bond between a particular output of the stimulus identification process and its assigned response counter that produces the difference in the magnitude of the relative frequency effect observed between high and low compatibility ELA conditions in the present experiment.

REFERENCES

Berlyne, D. E. Conflict and choice time. *British Journal of Psychology*, 1957, **48**, 106–118.

Bertelson, P., & Tisseyre, G. Choice reaction time as a function of stimulus versus response relative frequency of occurrence. *Nature*, 1966, **212**, 1069–1070.

Brainard, R. W., Irby, T. S., Fitts, P. M., & Alluisi, E. A. Some variables influencing the rate of gain of information. *Journal of Experimental Psychology*, 1962, **63**, 105–110.

Fitts, P. M., Peterson, J. R., & Wolpe, G. Cognitive aspects of information processing: II. Adjutments to stimulus redundancy. *Journal of Experimental Psychology*, 1963, **65**, 423–432.

Hawkins, H. L., & Hosking, K. Stimulus probability as a determinant of discrete choice reaction time. *Journal of Experimental Psychology*, 1969, **82**, 435–440.

Hawkins, H. L., Thomas, G. B., & Drury, K. B. Perceptual versus response bias in discrete reaction time. *Journal of Experimental Psychology*, 1970, **84**, 514–517.

Hyman, R. Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, 1953, **45**, 188–196.

Keele, S. W. An analysis of S-R compatibility. Unpublished doctoral dissertation, University of Wisconsin, 1966.

Leont'ev, A. N., & Krinchik, C. P. On certain aspects of human information processing. *Soviet Psychology & Psychiatry*, 1963, **1**, 24–31.

Posner, M. L. Components of skilled performance. *Science*, 1966, **152**, 1712–1718.

Smith, E. E. Choice reaction time: an analysis of the major theoretical positions. *Psychological Bulletin*, 1968, **69**, 77–110.

Stone, M. Models for choice reaction time. *Psychometrika*, 1960, **25**, 251–260.

(Received May 28, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 287–294

AVERAGING OF MOTOR MOVEMENTS: TESTS OF AN ADDITIVE MODEL

IRWIN P. LEVIN,[1] JOHN L. CRAFT,[2] AND KENT L. NORMAN

University of Iowa

The *S*s were required to make several successive blind linear movements of controlled length and then had to make an additional uncontrolled blind linear movement to represent the average of the several fixed movements (stimuli). Support was found for a class of additive models. With two movements to be averaged, the responses could be approximated by a simple average of the stimulus scale values, where the subjective scale values of the longer fixed movements were estimated to be less than the objective lengths. For conditions with three or five stimuli, the averaging responses could be approximated by a weighted average of the stimulus scale values, where the weight or influence of the last fixed movement was estimated to be slightly greater than the weights of the earlier fixed movements. Results were discussed in terms of the relationship between retention and serial integration processes.

Two experiments are presented which use the following task: *S* makes several successive controlled motor movements—moving a sliding block a fixed distance across a horizontal plane while blindfolded—and then is required to make an additional (uncontrolled) blind linear movement to represent the *average* of the several fixed movements. The manner in which *S* integrates the successive fixed movements to arrive at his average was examined by assessing the effects of variables such as the number and sequencing of stimuli (fixed movements) within the context of a family of additive models of the integration process.

In the general form of the additive model, the integrative judgment is considered to be a weighted sum or average of the scale values of the stimuli, where the weight of a given stimulus is assumed to depend on its serial position independent of the value of the stimulus. For the present task, a weighted average model would take the following form:

$$R_A = \Sigma w_i s_i, \qquad [1]$$

where R_A is the averaging response, s_i and w_i are the subjective scale value and weight, respectively, of the stimulus (fixed movement) at Ordinal Position i, and $\Sigma w_i = 1$. If the values of the stimuli at each ordinal position are varied in a factorial design, it can be shown that an additive model predicts no interactions in the analysis of variance tests (see Anderson, 1964a). The present experiments employed factorial designs, and the interaction tests were used to assess the goodness of fit of the models.

While much of the earlier work with these models has dealt with verbal stimuli, serial integration tasks with psychophysical stimuli such as line length, loudness of tone, and lifted weight have been investigated by Anderson and his associates (Anderson, 1967; Parducci, Thaler, & Anderson, 1968; Weiss & Anderson, 1969). Anderson (1970) reports mixed support for an additive model based on averaging theory and suggests that further study is warranted. The present task is related to these earlier tasks and could serve to increase the generality of the findings and model development in this area. However, the present task has two features which, in combination, distinguish it from previous tasks: (*a*) Stimulus magnitude is manipulated along a continuous dimension, and responses are executed and recorded along the *same* dimension; and (*b*) visual as well as verbal stimuli are eliminated. Craft and Hinrichs (1971), in a retention study conducted with essentially the same apparatus and procedures of the

[1] Requests for reprints should be sent to Irwin P. Levin, Department of Psychology, University of Iowa, Iowa City, Iowa 52240.

[2] Now at Wartburg College, Waverly, Iowa 50677.

present studies, showed that recall performance as a function of stimulus sequencing was not affected by the opportunity for selective attention (precuing S as to which stimulus he would have to reproduce). In general, this would not be predicted for visual or verbal stimuli (Posner, 1967). If mediating processes such as attention are analogous in retention and serial integration, then the present task allows for the study of serial integration where these processes may play a reduced role.

Experiment I

In Exp. I, S was asked to produce the average of two successive fixed linear movements. From the additivity assumption of the model, i.e., that w_1 and w_2 are invariant across values of s_1 and s_2, it is predicted that an analysis of variance on the averaging responses should reveal no interaction between stimulus value at Ordinal Position 1 and stimulus value at Ordinal Position 2. The interval between the two fixed movements and the interval between the second fixed movement and the averaging response were varied. Additional tests of alternative forms of Equation 1 depend on the relationship between the w values and these intervals. If w_1 and w_2 vary as a function of the time intervals, then there are expected to be interactions involving the time variables.

Method

Design.—Each of the two fixed movements (S_1, S_2) on a given trial was of length 15, 35, or 55 cm. The interval between the two fixed movements (T_1) was 5 or 10 sec., and the interval between the completion of the second fixed movement and the required averaging response (T_2) was 5 or 10 sec. A $3 \times 3 \times 2 \times 2$ factorial design was employed with each S receiving all 36 combinations of S_1, S_2, T_1, and T_2.

Subjects.—The Ss were 24 (14 male, 10 female) students in introductory psychology courses at the University of Iowa.

Apparatus.—Two free-moving, 5-cm.-square Plexiglas blocks were mounted on a wooden meter stick. The right-hand block was locked in place by E when defining the length of the fixed movements and was moved out of the way when S was making his averaging response. The other block had a knob for fingertip grasping and was moved from left to right by S with his right hand to present the fixed movements and to indicate his averaging response. The apparatus was placed on a table parallel to the coronal plane. The S was blindfolded to insure that his responding was not based on visual cues.

Procedure.—Each S received a different random order of presentation of the 36 experimental trials. On each trial, S was required to complete the fixed movements S_1 and S_2 (time to complete a movement and return to the starting position was approximately 8 sec.) with designated time intervals T_1 and T_2 and was then required to move the sliding block a distance which he estimated to be the average of the two fixed movements. Responses were recorded to the nearest millimeter. The S heard the following tape-recorded sequence of commands: "engage," "move one," "return," "move two," "return," "average," and "return and disengage." Upon hearing the command "engage," S grasped the knob and waited for the first "move" command. At the "move" command, S was instructed to "smoothly move your block down the slide, to your right, until your movement is stopped by the other block." The S held his block at the stop until the command "return" was presented. The S then returned his block to the starting position. Following the second "move" and "return" commands, S heard "average," which "will be a signal to estimate the average of the lengths of the first two movements." At the "return and disengage" command, S returned his block and removed his hand from the apparatus until the next "engage" command. The interval between the "disengage" command on one trial and the "engage" command starting the next trial was 10 sec. The S was not permitted to remove his hand from the apparatus during a trial and was given no feedback concerning the accuracy of his averaging response. Prior to the experimental trials, four practice trials were given. These consisted of each combination of T_1 and T_2, with fixed movement lengths more extreme than the values of S_1 and S_2 used on the experimental trials. A 3-min. rest interval was inserted midway through the session.

Results and Discussion

The mean averaging response for each condition is shown in Table 1. It can be seen that in most cases, Ss underestimated the actual mean of the two fixed movements. If Ss' averaging responses corresponded exactly to the physical means, then the values 25, 35, and 45 would have been obtained for the successive levels of S_1 and S_2. For S_1, the corresponding mean values were 24.70, 32.98, and 43.40; for S_2, the mean values were 24.85, 33.18, and 43.05. An analysis of variance performed on these data showed that the main effects of value of S_1 and value of S_2 were both highly significant, F (2, 46) = 875.51 and

1,100.35, respectively. The time intervals T_1 and T_2 were not significant sources of variance, but the latter effect approached significance at the .05 level, with the mean averaging response being slightly less following a 10-sec. stimulus–response interval than following a 5-sec. interval. None of the interactions involving T_1 or T_2 approached statistical significance ($p > .20$ in each case).

Contrary to the predictions of the model, there was a statistically significant interaction between the stimulus values S_1 and S_2, F (4, 92) = 7.13, $p < .01$. This interaction is graphed in Fig. 1, along with a plot of the physical means for each of the nine $S_1 \times S_2$ combinations. In can be seen that the lines are very nearly parallel. In this regard, it should be noted that the proportion of the variance contributed by the interaction term was only about one-hundredth of the proportion contributed by either main effect. To further assess the reliability of the interaction, an analysis of variance test was performed for each *S* individually by treating the four $T_1 \times T_2$ time interval combinations as replications

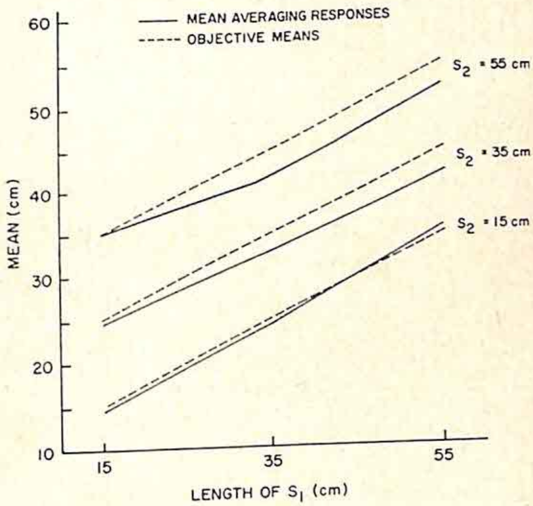

FIG. 1. Mean averaging response as a function of S_1 and S_2 length in Exp. I.

TABLE 1

MEAN AVERAGING RESPONSE FOR EACH EXPERIMENTAL CONDITION IN EXPERIMENT I

Time intervals (sec.) T_1–T_2	Value of S_2 (cm.)	Value of S_1 (cm.) 55	35	15	*M*
5–5	55	51.4	42.2	35.5	43.1
	35	42.4	33.6	25.2	33.7
	15	36.8	23.9	15.3	25.3
	M	43.5	33.2	25.3	34.0
5–10	55	52.6	41.8	34.4	42.9
	35	41.5	32.6	24.0	32.7
	15	35.5	24.3	14.2	24.7
	M	43.2	32.9	24.2	33.4
10–5	55	52.8	42.9	35.0	43.6
	35	42.2	32.5	24.9	33.2
	15	34.7	24.4	14.4	24.5
	M	43.2	33.3	24.8	33.8
10–10	55	52.2	40.8	34.9	42.6
	35	42.8	32.3	24.2	33.1
	15	36.0	24.4	14.4	24.9
	M	43.7	32.5	24.5	33.5

Note.—Mean averaging responses are measured in centimeters.

to obtain the necessary error terms. For each of the 24 *S*s, the main effects of value of S_1 and value of S_2 were both highly significant (df = 2, 6, $p < .001$, in each case). On the other hand, the interaction $S_1 \times S_2$ (df = 4, 12) was significant at or beyond the .05 level for only 6 of the *S*s. Thus, for 75% of the *S*s, no significant interaction was found even though large main effects were obtained. Consequently, the small but statistically significant interaction obtained with group data will be interpreted as being a fortuitous consequence of the extremely high power inherent in the group analysis. Further additivity tests will be included in Exp. II, but tentative support will be assumed for an averaging model.

Alternative forms of the averaging model can now be assessed. Since none of the interactions involving T_1 and T_2 approached significance, the weighting parameters w_1 and w_2 are assumed not to depend on the time intervals. The weighting parameters could vary as a function of ordinal position. However, a comparison of different stimulus sequences with the same pair of stimulus values (e.g., S_1 = 55, S_2 = 35; and S_1 = 35, S_2 = 55) reveals no systematic order effect. Estimates of w_1 and w_2 were obtained through functional measurement techniques (see Anderson, 1970).

TABLE 2

MEAN AVERAGING RESPONSE AND PREDICTED VALUE FOR EACH COMBINATION OF S_1 AND S_2 IN EXPERIMENT I

Value of S_2 (cm.)	Value of S_1 (cm.)			
	15	35	55	M
15	14.6 (15.9)	24.2 (24.2)	35.7 (34.3)	24.8
35	24.6 (24.2)	32.8 (32.5)	42.2 (42.6)	33.2
55	35.0 (34.3)	41.9 (42.6)	52.3 (52.8)	43.1
M	24.7	33.0	43.4	33.7

Note.—Predicted values are in parentheses.

The resulting estimates were $w_1 = .51$, $w_2 = .49$, indicating only a slight tendency toward primacy. A subsequent analysis revealed no statistically significant order effect, $F < 1$. A parsimonious interpretation is that the averaging response equals the simple arithmetic mean of the subjective scale values s_1 and s_2.

To obtain estimates of the scale values, functional measurement techniques were used assuming a simple average model and assuming that $s_1 = s_2$ for a given fixed movement length. For Lengths 15, 35, and 55, the corresponding scale value estimates were 15.86, 32.46, and 52.76. The predictions of a simple average model using these parameter estimates can be compared with the observed means in Table 2 for each $S_1 \times S_2$ combination, averaged over values of the time intervals. Most of the observed values are within 2% of the predicted values. The mean discrepancy between predicted and observed values is 2.3%. In general, then, the averaging response can be approximated by the simple average of the subjective scale values, where the subjective scale values of the longer fixed movements are less than the corresponding objective values.

EXPERIMENT II

Several aspects of the results of Exp. I appear to warrant further study. First, the statistically significant $S_1 \times S_2$ interaction, while slight in magnitude and not apparently systematic in nature, does cast some doubt on the applicability of an additive model. The additivity assumption was explored in more depth in Exp. II, in which there were more than two fixed movements to be averaged. For each of two experimental groups, stimulus value at each of three successive ordinal positions was varied factorially in a 3^3 design, allowing four independent additivity (interaction) tests. For a third group, stimulus value at each of five successive ordinal positions was varied in a 2^5 design, allowing 26 additivity tests.

The use of a three- or five-stimulus serial integration task permits an expanded analysis of the serial position effect. An additional manipulation that could affect the direction and magnitude of the order effect in Exp. II is the required mode of responding. For the three-stimulus condition, some *S*s were required to estimate a running average following each successive fixed movement in the sequence; other *S*s were required to average only at the end of the sequence. Previous research with other serial integration tasks such as personality impression formation (Anderson, 1965; Levin & Schmidt, 1969; Stewart, 1965) and number averaging (Anderson, 1964b; Hendrick & Costantini, 1970) found primacy effects with end-only responding and recency effects with continuous responding. Anderson and Hubert (1963) suggest that primacy is primarily caused by decreased attention to later stimuli but that primacy is destroyed under conditions which cause *S* to attend more completely to the later stimuli. This account thus appeals to attentional factors underlying order effects. To the extent that such mediating processes are reduced in the present motor task, one would expect different order effects than found in the earlier studies.

Method

Three-stimulus condition.—The *S*s were 24 (12 males, 12 females) students in introductory psychology courses at the University of Iowa. The apparatus was the same as that used in Exp. I. Six males and six females were randomly assigned to each of two groups. The *S*s in the end-only-average (EOA)

group executed three fixed movements (S_1, S_2, S_3) on each trial and then were required to move the sliding block a distance which they estimated to be the average of the three fixed movements. The *S*s in the running-average (RA) group also executed three fixed movements on a trial but were required to respond after each fixed movement. The first response on a trial was simply an attempt to reproduce the length of S_1; the second and third responses were estimated averages of S_1 and S_2, and S_1, S_2, and S_3, respectively. Each of the fixed movements on each trial was of Length 15, 35, or 55 cm. for both groups. Within each group, a 3^3 factorial design was employed with each *S* receiving all combinations of S_1, S_2, and S_3. Each *S* received a different random order of presentation of the 27 experimental trials.

For both groups, the time to complete a fixed movement and return to the starting position was 8 sec. The time to complete an averaging response and return to the starting position was 10 sec. Group EOA sat quietly during 10-sec. intervals between S_1–S_2 and S_2–S_3, whereas Group RA made averaging responses during these two intervals. The interval between the completion of a fixed movement and the required averaging response was 3 sec. Prior to the experimental trials, three practice trials were given with lengths of S_1, S_2, and S_3 different from those used on the experimental trials.

Five-stimulus condition.—The *S*s were 12 (6 males, 6 females) students who received essentially the same procedure as *S*s in Group EOA above, except that they executed five rather than three successive fixed movements (S_1, S_2, S_3, S_4, S_5) to be averaged at the end of each trial, and they did not have 10-sec. interstimulus intervals. The fixed movement at each ordinal position was of Length 15, or 45 cm., forming a 2^5 factorial design. Each *S* received a different random order of presentation of the 32 experimental stimulus sequences following four practice trials.

Results and Discussion

Three-stimulus condition.—The mean averaging response of the three successive fixed movements is shown in Table 3 for each condition in each group. As in Exp. I, most of the cells reveal an underestimation of the actual mean of the three fixed movements. Analysis of variance performed with groups as a factor revealed no significant main effect of groups or interactions between groups and the other factors of the design. However, to provide independent tests of the model, Groups RA and EOA were analyzed separately. The main effects of value of S_1, value of S_2, and value of S_3 were highly significant for each group, F (2, 22) > 100 in each case.

In testing the additivity assumptions of the model applied to the averaging of three successive fixed movements, there are four interaction tests for each group. For Group RA, none of the interaction terms were

TABLE 3

MEAN AVERAGING RESPONSE AND PREDICTED VALUE FOR EACH COMBINATION OF S_1, S_2, AND S_3

Group and value of S_2	S_1 = 15				S_1 = 35				S_1 = 55			
	Value of S_3				Value of S_3				Value of S_3			
	15	35	55	*M*	15	35	55	*M*	15	35	55	*M*
RA												
15	15.9 (16.3)	22.6 (22.4)	32.3 (30.5)	23.6	20.5 (20.9)	27.0 (27.1)	34.2 (35.1)	27.2	26.3 (26.9)	34.8 (33.1)	39.7 (41.2)	33.6
35	20.0 (20.9)	28.3 (27.1)	34.7 (35.1)	27.7	25.0 (25.5)	33.1 (31.7)	39.9 (39.8)	32.7	31.1 (31.6)	37.3 (37.7)	46.5 (45.6)	38.3
55	28.9 (26.9)	31.5 (33.1)	41.8 (41.2)	34.1	31.9 (31.6)	36.9 (37.7)	44.5 (45.8)	37.8	36.0 (37.6)	43.1 (43.8)	53.8 (51.8)	44.3
M	21.6	27.5	36.3	28.5	25.8	32.3	39.5	32.6	31.1	38.4	46.7	38.7
EOA												
15	14.2 (17.1)	24.7 (22.8)	32.1 (30.5)	23.7	22.0 (22.4)	27.2 (28.1)	37.1 (35.8)	28.8	29.8 (29.4)	35.2 (35.1)	42.1 (42.8)	35.7
35	20.3 (21.6)	25.7 (27.3)	34.0 (35.0)	26.7	25.8 (26.8)	32.0 (32.6)	41.1 (40.2)	33.0	35.4 (33.9)	39.8 (39.6)	48.9 (47.3)	41.4
55	29.9 (27.6)	33.5 (33.3)	40.9 (41.0)	34.8	35.7 (32.0)	39.7 (38.6)	45.5 (46.3)	40.3	39.1 (39.9)	44.0 (45.7)	51.4 (53.3)	44.8
M	21.5	28.0	35.7	28.4	27.8	33.0	41.2	34.0	34.8	39.7	47.5	40.6

Note.—Predicted values appear in parentheses.

statistically significant at the .05 level. (Additionally, the interaction tests for averaging responses at Ordinal Position 2 for Group RA were nonsignificant.) For Group EOA, there was one significant interaction; for the $S_1 \times S_2$ interaction, $F\ (4, 44) = 3.83$, $p < .01$. However, examination of the interaction revealed no systematic trend that could seriously question the additivity assumption. Thus further analysis within the context of a weighted average model was conducted.

Functional measurement techniques were used to estimate the weighting and scale value parameters for each group. For Group EOA, the resulting estimates were as follows: $w_1 = .34$, $w_2 = .29$, $w_3 = .37$; $s15 = 17.11$, $s35 = 32.56$, $s55 = 53.35$. In comparing the weighting parameters, it can be seen that the third stimulus receives the greatest weight in the averaging process. However, there is no orderly recency effect because the second stimulus is not weighted more than the first. The scale value estimates are similar to those of Exp. I, again suggesting that subjective scale values are less than the objective values for the longer fixed movements. For Group RA, the pattern of parameter estimates is similar to that of Group EOA. The following estimates were obtained: $w_1 = .29$, $w_2 = .30$, $w_3 = .41$; $s15 = 16.27$, $s35 = 31.69$, $s55 = 51.82$.

To further compare the serial position effects between Groups EOA and RA, the parameters w_1, w_2, and w_3 were estimated individually for each *S* in each group, and linear and quadratic trend components were computed. Groups EOA and RA did not differ significantly in mean linear trend component or mean quadratic trend component.

The predictions of a weighted average model using the group parameter estimates can be compared with the observed means in Table 3. Most of the discrepancies are quite small. For Group EOA, the mean discrepancy between predicted and observed values is 3.9%. For Group RA, the corresponding value is 2.7%. The better fit of the model in Group RA than in Group EOA is not surprising in light of the fact that a significant interaction was found for Group EOA but not for Group RA.

In an analogous fashion, an averaging model was applied to responses at Ordinal Position 2 for Group RA. The results were quite similar to those of Exp. I, which involved the same stimulus combinations but in which *S*s gave an end-only averaging response. As in Exp. I, the weighting parameter estimates were nearly equal ($w_1 = .49$, $w_2 = .51$) and a simple average model was assumed. The scale value estimates were $s15 = 14.95$, $s35 = 32.72$, $s55 = 52.40$. The mean discrepancy between predicted and observed values was only 1.6%.

Five-stimulus condition.—The mean averaging response of the five successive fixed movements is shown in Table 4 for each condition along with the physical mean for that condition. As before, most of the cells show an underestimation. Analysis of variance revealed that the main effects of the values of S_1, S_2, S_3, S_4, and S_5 were all highly significant, $F\ (1, 11) > 40$ in each case.

In testing the additivity assumptions of the weighted average model, 3 out of the 26 interactions were significant ($.01 < p < .05$ in each case). These were the three-way interactions $S_1 \times S_2 \times S_3$, $S_1 \times S_4 \times S_5$, and $S_3 \times S_4 \times S_5$. A plot of these interactions did not suggest any systematic discrepancy from the additivity assumption, and the average proportion of the variance contributed by these interactions was only about one-fortieth of the average proportion of the variance contributed by the main effects. None of the other interaction terms approach significance at the .05 level.

The weighting and scale value parameters were estimated to be as follows: $w_1 = .17$, $w_2 = .22$, $w_3 = .18$, $w_4 = .19$, $w_5 = .25$; $s15 = 16.19$, $s45 = 40.16$. The pattern of serial position weights can be seen to be similar to those of the three-stimulus conditions. The last stimulus receives the greatest weight, but there is no orderly recency effect.

The predictions of the model using the above parameter estimates can be com-

TABLE 4

MEAN AVERAGING RESPONSE AND PREDICTED VALUE FOR EACH COMBINATION OF S_1, S_2, S_3, S_4, AND S_5

S_1	S_2	S_3	S_4	S_5	Physical M (cm.)	Averaging response	Predicted value
15	15	15	15	15	15	13.9	16.2
15	15	15	15	45	21	21.8	22.2
15	15	15	45	15	21	22.4	20.8
15	15	15	45	45	27	25.5	26.8
15	15	45	15	15	21	20.8	20.5
15	15	45	15	45	27	26.2	26.5
15	15	45	45	15	27	25.9	25.1
15	15	45	45	45	33	30.7	31.1
15	45	15	15	15	21	22.8	21.2
15	45	15	15	45	27	27.7	27.2
15	45	15	45	15	27	26.4	25.8
15	45	15	45	45	33	31.1	31.8
15	45	45	15	15	27	25.9	25.6
15	45	45	15	45	33	31.8	31.6
15	45	45	45	15	33	28.9	30.1
15	45	45	45	45	39	36.9	36.1
45	15	15	15	15	21	20.0	20.3
45	15	15	15	45	27	28.3	26.3
45	15	15	45	15	27	25.4	24.8
45	15	15	45	45	33	31.7	30.8
45	15	45	15	15	27	24.1	24.6
45	15	45	15	45	33	29.4	30.6
45	15	45	45	15	33	27.7	29.2
45	15	45	45	45	39	35.5	35.2
45	45	15	15	15	27	25.9	25.3
45	45	15	15	45	33	30.8	31.3
45	45	15	45	15	33	29.5	29.9
45	45	15	45	45	39	34.0	35.9
45	45	45	15	15	33	29.6	29.6
45	45	45	15	45	39	35.3	35.6
45	45	45	45	15	39	32.6	34.2
45	45	45	45	45	45	42.6	40.2

pared with the observed means in Table 4. Again, most of the discrepancies are quite small. The mean discrepancy between predicted and observed values is 3.3%. The largest discrepancies occur for sequences of all-15-cm. and all-45-cm. fixed movements. In Tables 2, 3, and 4 it can be seen that the scale value estimate of *s15* was always too high to accurately predict the averaging response for a stimulus sequence of all-15-cm. fixed movements. However, adjusting this estimate downward would have caused greater discrepancies in other cells. Thus it appears that the parameters are different when *S* is averaging a sequence of mixed stimulus values than when all the values are the same.

General Discussion

With *S* given the somewhat ambiguous instruction to "average," the responses suggest that *S* attempted to reproduce the mean of the several fixed movements and was relatively accurate in doing so. There was, however, an underestimation tendency and, in Exp. II, a serial position effect. These data could be accounted for reasonably well by a weighted average model. There were several departures from the model's prediction of no interactions, but the observed interactions were small in magnitude and appeared to be nonsystematic.

Estimates of serial position weights in Exp. II yielded order effects different from those found in studies using serial integration tasks such as impression formation or number averaging. The order effects found in these earlier studies—typically, primacy with end-only responding and recency with continuous responding—have been attributed to attentional factors (Anderson & Hubert, 1963). In the present three-stimulus conditions, there were no significant differences in trend components for end-only and continuous responding (running average). For three- and five-stimulus conditions, there was no systematic primacy or recency, but there was a slightly enhanced effect of the last stimulus. One possible interpretation of the difference between the results with the present task and with earlier tasks is that attentional processes may be different. It has already been suggested that such mediating processes play a reduced role in the retention of kinesthetic as compared to visual and verbal stimuli (Craft & Hinrichs, 1971; Posner, 1967). This could also be true of serial integration of motor responses.

Functional measurement techniques used in the present study yielded estimates of subjective scale values of the longer fixed movements which were less than the objective values. This is consistent with the findings of stimulus trace decays in motor retention (Adams & Dijkstra, 1966; Stelmach, 1969) and supports a commonality of processes operative in motor retention and serial integration.

Several authors have suggested that integration processes affect motor retention. According to Pepper and Herman (1970), when *S* has to recall a standard motor response he reproduces a trace of intensity intermediate in value to the standard and interpolated inter-

fering intensities. Craft (1971) found support for a formal model of motor short-term memory in which interference of one motor movement with recall of another is assumed to take place in the form of memory trace alteration or partial integration. In the future, studies of motor memory and studies of motor serial integration could serve to complement each other.

REFERENCES

Adams, J. A., & Dijkstra, S. Short-term memory for motor responses. *Journal of Experimental Psychology*, 1966, **71**, 314–318.

Anderson, N. H. Note on weighted sum and linear operator models. *Psychonomic Science*, 1964, **1**, 189–190. (a)

Anderson, N. H. Test of a model for number-averaging behavior. *Psychonomic Science*, 1964, **1**, 191–192. (b)

Anderson, N. H. Primacy effects in personality impression formation using a generalized order effect paradigm. *Journal of Personality and Social Psychology*, 1965, **2**, 1–9.

Anderson, N. H. Application of a weighted average model to a psychophysical averaging task. *Psychonomic Science*, 1967, **8**, 227–228.

Anderson, N. H. Functional measurement and psychophysical judgment. *Psychological Review*, 1970, **77**, 153–170.

Anderson, N. H., & Hubert, S. Effects of concomitant verbal recall on order effects in personality impression formation. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 379–391.

Craft, J. L. Memory trace interference and decay as factors influencing the short-term retention of motor responses. Unpublished doctoral dissertation, University of Iowa, 1971.

Craft, J. L., & Hinrichs, J. V. Short-term retention of simple motor responses: Similarity of prior and succeeding responses. *Journal of Experimental Psychology*, 1971, **87**, 297–302.

Hendrick, C., & Costantini, A. F. Number averaging behavior: A primacy effect. *Psychonomic Science*, 1970, **19**, 121–122.

Levin, I. P., & Schmidt, C. F. Sequential effects in impression formation with binary intermittent responding. *Journal of Experimental Psychology*, 1969, **79**, 283–287.

Parducci, A., Thaler, H., & Anderson, N. H. Stimulus averaging and the context for judgment. *Perception & Psychophysics*, 1968, **3**, 145–150.

Pepper, R. L., & Herman, L. M. Decay and interference effects in the short-term retention of a discrete motor act. *Journal of Experimental Psychology*, 1970, **83** (2, Pt. 2).

Posner, M. I. Characteristics of visual and kinesthetic memory codes. *Journal of Experimental Psychology*, 1967, **75**, 103–107.

Stelmach, G. E. Short-term motor retention as a function of response similarity. *Journal of Motor Behavior*, 1969, **1**, 37–45.

Stewart, R. H. Effect of continuous responding on the order effect in personality impression formation. *Journal of Personality and Social Psychology*, 1965, **1**, 161–165.

Weiss, D. J., & Anderson, N. H. Subjective averaging of length with serial presentation. *Journal of Experimental Psychology*, 1969, **82**, 52–63.

(Received May 29, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 295–298

FUNCTION OF INTERMEDIATE RESPONSES OF A BEHAVIOR CHAIN [1]

IN-MAO LIU [2] AND HSEN-HSING MA

National Taiwan University

With a chain of responses, the factor of overt or covert execution of intermediate steps was manipulated orthogonally to chain length. Quite differently from the case of a two-step response chain, it was found that in the case of a four-step response chain, covert execution of intermediate steps resulted in inferior performance in the early stage of training and more rapid improvement during the later stage. It was also found that response time increased linearly with chain length in the overt condition but nonlinearly in the covert condition. These findings point out the ways by which a cognitive process may be distinguished from a simple response chain.

Behavior usually consists of a chain of responses instead of a single momentary response. The nature of a response chain has long been an interesting object of study. Sometimes a response chain is observable overtly, sometimes only covertly. A covert response chain is a sort of cognitive process like scanning, thinking, or problem solving, where it is hard for an *O* to tell what are the intermediate responses. The present study attempts to study the function of the intermediate responses of a behavior chain.

For the present study, let us consider eight buttons on a board. Under one condition, *S* will be instructed to press each of two or four buttons according to some rule. Under another condition, *S* will be required to press only the last button according to the same rule without pressing the intermediate buttons. The former condition is typical of an observable chain of acts; the latter is typical of an implicit searching or problem-solving task. When only two buttons are involved, it is a short response chain. When four buttons are involved, it is a long response chain. By manipulating the length of a response chain, the present study investigates the function of intermediate responses.

If a response chain is short, in comparison to the situation where overt execution of the intermediate steps is prohibited, execution of every intermediate step may lengthen the total time required to complete the whole response chain. However, if a response chain is long, omission of the intermediate steps may result in poorer performance. Because every step is based on the immediately preceding step as a reference point, the location of each of these reference points should be less distinct in the case of a long covert chain. This is as if the multiplication of two three-digit numbers without writing down the intermediate steps seems more difficult than carrying out the multiplication step by step. The investigation of this type of problem may throw light on the linkage between simple behavior and complex cognitive process.

METHOD

Subjects.—The *S*s were 126 freshmen from introductory psychology courses. They participated in the experiment to fulfill a course requirement.

Apparatus and task.—The main part of the apparatus consisted of a device for stimulus presentation and a response panel on which eight buttons arranged as a column were mounted. The distance between two adjacent buttons was 3.5 cm. The stimulus consisted of a card on which a column of eight dots were printed, one of which had a red circle around it to indicate it as the starting point. The stimulus card was presented in the left visual field. The response panel was located in the right.

There were three different tasks. For one task (replication task), depending on which dot had a circle around it, *S* was to press the corresponding button on the response panel. For instance, if the second dot was identified as the starting point, *S* was to press the second button on the response

[1] This study was supported by a grant from the National Science Council.

[2] Requests for reprints should be sent to In-mao Liu, Department of Psychology, National Taiwan University, Taipei, Taiwan 107, Republic of China.

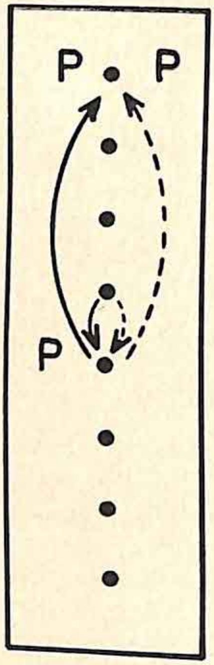

FIG. 1. Response panel with a column of eight buttons. (In these examples, the supposed starting point as indicated in the stimulus card is the fourth button. The solid line indicates the two-step overt response chain; the dotted line, the two-step covert response chain. The letter P indicates the button S is required to press.)

panel. For another task (symmetry task), after identifying the starting point from the stimulus card, S was to press the button symmetrical to the indicated one with respect to the axis bisecting the column of buttons into the upper and lower halves. For example, if the second dot of the stimulus card had a circle around it, S was to press the second from the last (i.e., seventh) button. For still another task (multiplication task), S was to press the button having twice the distance of the botton indicated. The distance of the indicated button was always measured from the nearer edge (upper or lower) of the response panel. The distance between the top button (or bottom button) and the upper edge (or lower edge) of the response panel was equal to the interbutton distance. For instance, if the second dot of the stimulus card was identified as the starting point, S was to press the fourth button; if the eighth dot was identified as the starting point, he was to press the seventh button. Since, in the latter case, if the distance between the eighth button and the lower edge, which is the nearer edge in this case, was one unit, then the distance between the seventh button and the same edge was two units.

Procedure.—The Ss were assigned to one of seven groups of 18 Ss each, in the order of their appearance. Four experimental groups were used to form a 2×2 factorial design, in which the length of response chain and the presence or absence of executing the intermediate steps were manipulated. There were two lengths of response chain: a two-step and a four-step response chain. The two-step response chain consisted of the symmetry task followed by the multiplication task; the four-step response chain consisted of the symmetry, multiplication, symmetry, and multiplication tasks in succession. For the overt condition of executing the intermediate steps, S was required to press a button for each component task; for the covert condition of omitting the intermediate steps, to press the button for the final component task only. Examples of two-step overt and covert response chains are shown in Fig. 1. With E's ready signal, S rested his right index finger on a telegraph key on the right side of the response panel. When the card was exposed, a clock started at the same time. When S pressed the final button correctly, a small bulb was lit and the clock (invisible to S) stopped automatically. Otherwise, he was to start again. The stimulus was exposed until S pressed the final correction button. Since the third or sixth dot as a starting point resulted in pressing the same third or sixth button as correct, they were not used as starting points. On every trial, one of the remaining six dots served as a starting point. For every block of 6 trials, the six dots were randomly used as the starting points, each dot appearing as the starting point once. According to the preliminary experiment, after about 10 blocks of trials, Ss of the covert groups started to take advantage of six invariant relationships between starting points and ending points by disregarding the intermediate steps entirely. Therefore, in the present experiment, each S was given 48 trials, i.e., 8 blocks of 6 trials.

For the condition of covert response chain, S was required to rest his finger on the telegraph key until the time of pressing the final correct button. There was a signal bulb on the side of E (invisible to S) such that when S lifted his finger from the key, it was lit. This device was used since, otherwise, S tended to move his finger to help carry out the covert response chain.

There were three control groups. For one control group, the replication task alone was given. For another control group, the symmetry task alone was given. For still another control group, the multiplication task alone was given. The same six dots served as the starting points as in the experimental groups. Each S was also given 48 massed trials.

RESULTS

Figure 2 shows the improvement of performance in the course of training for the four experimental groups. The long response chain took more time to execute than the short response chain. However, execution of the intermediate steps or its omission affected the performance differently, depending on the length of a re-

sponse chain. For the short response chain, omission of executing the intermediate steps tended to produce superior performance, although this tendency was very small. For the long response chain, omission of executing the intermediate steps resulted in inferior performance. In the latter case, the magnitude of the performance difference was initially large, but decreased in the course of training. An analysis of variance shows that training, chain length, and the factor of overt–covert execution of intermediate steps were significant sources of variance: F (7, 476) = 166.00, $p < .001$; F (1, 68) = 109.26, $p < .001$; and F (1, 68) = 8.82, $p < .01$, respectively.

As noted above, the Chain Length × Intermediate Steps interaction was significant, F (1, 68) = 11.55, $p < .01$. The interactions of Chain Length × Training and Intermediate Steps × Training were also significant, F (7, 476) = 15.65, $p < .001$, and F (7, 476) = 2.3, $p < .05$, respectively. According to Fig. 2, the latter effect was largely due to the greater improvement in performance for the covert condition in the case of the long response chain. The three-way interaction was significant, F (7, 476) = 3.24, $p < .01$.

Further analyses showed that as far as the first trial block is concerned, overt execution of the intermediate steps did not result in significantly superior performance, $t(34) = .42$, for the short response chain, but did result in inferior performance for the long response chain, $t(34) = 2.80$, $p < .001$. For the last trial block, no significant difference was obtained in either case, $t(34) = 1.29$, and $t(34) = .95$, respectively.

The addition of response time as a consequence of increasing chain length was analyzed in a different way. For the overt condition, the response time of each trial block for each S trained with the two-step response chain was doubled and this measure was compared with the response time of the four-step response chain. This is a rough comparison from which we can judge whether there are additional processes involved in the long response chain under

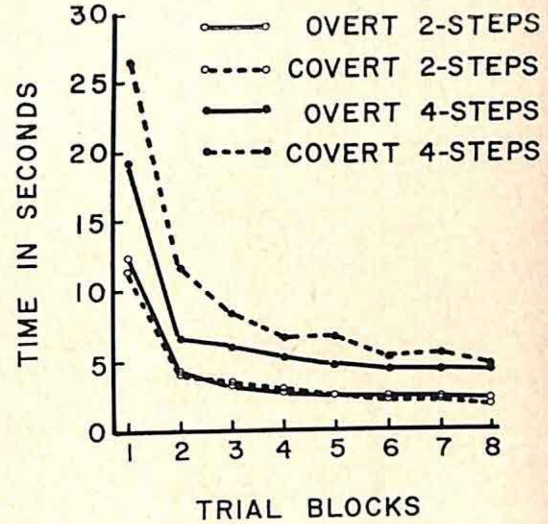

FIG. 2. Performance curves of the experimental groups.

the same overt condition. An analysis of variance shows that there was no significant difference, F (1, 34) = 1.50. In the covert condition, by contrast, even the doubled response time of the two-step was significantly shorter than that of the four-step group, F (1, 34) = 5.54, $p < .05$.

The data obtained from the three control groups give us information about the component mechanisms of a response chain. Figure 3 shows the performance curves for these control groups. It is obvious from this figure that the multiplication task is the most difficult one. The symmetry task seems slightly more difficult than the replication task. An analysis of variance showed that three tasks produced different performance levels, F (2, 51) = 22.39, $p < .001$. The training effect as well as the Treatments × Training interaction was significant, F (7, 345) = 47.89, $p < .001$, and F (14, 357) = 13.57, $p < .001$, respectively. Further analysis showed that the response time required for the multiplication task was significantly longer than that for the replication task and also longer than that for the symmetry task, t (51) = 6.37, $p < .001$ and t (51) = 5.06, $p < .001$, respectively. However, the response time required for the symmetry task did not differ significantly from that required for the replication task, t (51) = 1.31.

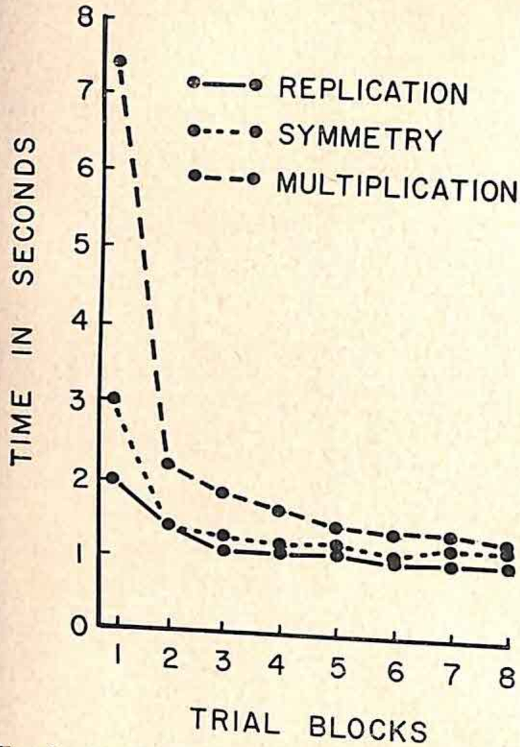

FIG. 3. Performance curves of the control groups.

DISCUSSION

The explanation for the fact that omission of the intermediate steps in the long-chain condition resulted in poorer performance may be as follows. When carrying out a chain of responses with the present task, each response should depend on the immediately preceding response. The immediately preceding response had to provide a reference point or starting point from which the next response could be executed. When the intermediate responses of a chain could be carried out only covertly, *S* had to retain in memory each immediately preceding reference point as well as the information about its relative position in the space. When the intermediate steps could be executed overtly, such retention of each reference point was not demanding, consequently resulting in less confusion. Presumably, the same mechanism underlies our thought process. Chess players are familiar with the fact that a novice usually cannot think more than one or two moves ahead without making overt trial moves.

Nevertheless, there is an advantage in the covert condition. Since no intermediate responses are required, we may guess that there is a lot of room for improvement in terms of the time required to carry out the whole response chain. This conjecture is in agreement with a larger improvement obtained in the covert condition with the long response chain. Common observations show that an expert chess player can think more than several moves ahead, probably because of his long-term training.

When execution of the intermediate steps was required, the time required to complete a response chain is fairly linear with chain length. On the other hand, the time required to carry out the whole response chain with covert intermediate steps increases more rapidly as chain length is increased. Taking the eighth block of trials, for the one-step chain, the average response time was 1.23 sec. (obtained by averaging those of the symmetry and multiplication tasks); for the two-step covert chain, 1.84 sec.; and for the four-step covert chain, 4.80 sec. This nonlinearity or nonadditivity indicates that a covert response chain may involve disproportionate increase in repeated attempts (covert) and repeated confirmations (covert) as the number of intermediate steps increases. Although in a different situation, a similar observation was made in an experiment by Archer, Bourne, and Brown (1955). These investigators studied the rate of concept recognition as a function of the number of irrelevant dimensions involved if the number of relevant dimensions is held constant. They found that as the number of irrelevant dimensions increases, time to solve increases also in a positively accelerated fashion. They suggested that *S* forgets which particular hypothesis he has tested in the past and proceeds to test them repeatedly.

In the case of the response chains with overt intermediate steps, it was pointed out that doubling the response time of the two-step chain did not result in a significant difference from the response time of the four-step chain. A slight difference could be due to doubling the component of identification time in the total response time, because, *S* had to identify the starting point before carrying out the response chain.

REFERENCE

ARCHER, E. J., BOURNE, L. E., JR., & BROWN, F. G. Concept identification as a function of irrelevant information and instructions. *Journal of Experimental Psychology*, 1955, **49**, 153–164.

(Received June 24, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 299–304

PREDICTION OUTCOME, S–R COMPATIBILITY, AND CHOICE REACTION TIME [1]

CHARLES P. WHITMAN AND E. SCOTT GELLER [2]

Virginia Polytechnic Institute and State University

Intrasequential effects of prediction outcome (PO) on choice reaction time (choice RT) were studied in a two-alternative reaction task under two levels of S–R compatibility. Given a correctly predicted event, choice RT was significantly shorter when the preceding PO was correct than when it was incorrect. For incorrectly predicted events in a Compatible condition, preceding correct POs lengthened choice RT but in an Incompatible condition, preceding correct POs shortened choice RT. For a Compatible condition the relationships between preceding PO and choice RT were consistent with continuous expectancy theory, but for an Incompatible condition the relationships were not explained by expectancy theory. The implication that an expectancy model best predicts the effects of independent variables on the stimulus identification component of the choice reaction process is discussed.

In a choice reaction time (RT) task, *S*s identify correctly predicted stimuli faster than incorrectly predicted stimuli (e.g., Bernstein & Reese, 1965; Geller, Whitman, Wrenn, & Shipley, 1971). In addition, the prediction outcome (PO) on trials immediately preceding a particular stimulus presentation has been shown to be a significant determinant of choice RT to that stimulus (Whitman & Geller, 1971). Specifically, Whitman and Geller demonstrated that choice RT to a particular stimulus was an inverse function of the number of consecutive correct POs preceding that stimulus and was a direct function of the run length of incorrect POs preceding that stimulus.

To account for the observed effects of preceding PO, Whitman and Geller (1971) considered the continuous expectancy hypothesis proposed by Geller and Pitz (1970) that an increase in the degree of expectancy for a particular stimulus alternative augments both the facilitation to react to a correctly predicted stimulus and the inhibition to react to an incorrectly predicted alternative. Whitman and Geller (1971) hypothesized that a correct PO on a particular trial would increase *S*'s confidence in a subsequent prediction, and thus increase *S*'s degree of expectancy for the occurrence of the predicted stimulus. On the other hand, an incorrect PO was assumed to decrease confidence and thus decrease *S*'s degree of expectancy for the stimulus predicted. Thus, *S*'s degree of expectancy for the predicted stimulus would be a direct function of the number of consecutive correct predictions, and an inverse function of the number of consecutive incorrect predictions. Therefore, it was suggested that preceding correct POs would facilitate a response to a correctly predicted stimulus and inhibit a response to an incorrectly predicted stimulus.

The relationships found between choice RT and the number of consecutive correct or incorrect POs preceding a given stimulus presentation were only partially explained by continuous expectancy theory (Whitman & Geller, 1971). That is, expectancy theory correctly hypothesized the observation that reactions to correctly predicted stimuli immediately preceded by incorrect predictions were slower than reactions to correctly predicted stimuli preceded by correct predictions. However, the finding that reactions to incorrectly predicted stimuli preceded by correct predictions were faster than reactions to incorrectly predicted stimuli preceded by incorrect

[1] This research was facilitated by college gift equipment contributed by the Western Electric Company and the Chesapeake and Potomac Telephone company. Special thanks are due to Richard F. Wrenn for constructing the logic equipment.

[2] Requests for reprints should be sent to E. Scott Geller, Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061.

predictions was inconsistent with the proposed expectancy hypotheses. To explain this inconsistency, a post hoc explanation suggested that preceding PO influenced the response component of the choice reaction process rather than the stimulus identification component (cf. Sternberg, 1969; Welford, 1960). Consequently, the observation that choice RT was a decreasing function of the number of preceding correct predictions and an increasing function of the number of preceding incorrect predictions, regardless of the current PO, was interpreted as a result of response facilitation and response inhibition rather than changes in stimulus expectancy.

The results of recent studies have indicated that stimulus probability and prediction (two indicators of *S*s' expectancy in a choice RT task) influence primarily the stimulus identificaton component of the choice reaction process. That is, stimulus probability, rather than response probability, was found to account for the effects of relative frequency of stimulus occurrence on choice RT (Hawkins, Thomas, & Drury, 1970); and correct predictions of the stimulus rather than preparation to execute the correct response accounted for the finding that faster reactions occur to predicted than to nonpredicted stimuli (Hinrichs & Krainz, 1970). Moreover, it has been shown that as the S–R relationship is made more incompatible, perceptual factors have less influence on choice RT and response effects become more prominent determinants of choice RT (Schvaneveldt & Chase, 1969).

The present experiment was designed to determine the extent to which continuous expectancy theory accounts for the effects of preceding POs under two S–R compatibility levels. Compatibility was defined in terms of the naturalness of the pairing between the stimulus and the response alternatives rather than the similarity between the set of stimuli and responses (cf. Smith, 1968). This operational definition of S–R compatibility is identical to the compatibility distinction considered by other RT researchers (e.g., Biederman & Kaplan, 1970; Schaffer, 1965).

Method

Subjects.—Forty students (31 males, 9 females), none of whom had prior experience in a reaction time experiment, participated individually in the experimental sessions. All *S*s were enrolled in introductory psychology classes at Virginia Polytechnic Institute and State University, and received optional research credit for their participation.

Apparatus and procedure.—The stimuli consisted of the symbol "Ω" which was presented by one of two digital readouts, each $\frac{1}{2} \times 2$ in. and separated on a horizontal axis by 14 in. The left and right locations were verbally labeled "left" and "right," respectively. The *S*s viewed the stimuli through a one-way mirror from a distance of approximately 40 in. The experimental room was darkened and *S*s could not see through the mirror except when a readout was illuminated.

Throughout the 300 trials of each session, and controlled over consecutive blocks of 50, the left stimulus occurred on 70% of the trials and right occurred on the remaining 30%. Runs of the more probable left stimulus never exceeded a length of seven, and runs of the less probable right stimulus never exceeded three. In the task instructions, *S*s were not given any information about the structure of the stimulus list. However, *E* informed *S*s that the stimuli were prerecorded on punched paper tape and emphasized the importance of both speed and accuracy. An initial 10 practice trials were given during which *S*s' questions concerning the task were answered.

Each trial included the following events in the order listed: *S*s' verbal prediction, a brief warning buzzer, a uniformly random time interval of between $\frac{1}{2}$ and 2 sec., a stimulus presentation, and *S*s' choice reaction which turned off the stimulus. The latency between the stimulus presentation and *S*s' response was measured in milliseconds by a Hunter Klockcounter. Anticipatory responses and inappropriate choices were reported to *E* by indicator lights.

A choice response was made by pulling one of two microswitches which were positioned on spatially separated handles. In Group Compatible (15 males, 5 females), *S*s pulled the right switch to identify a right stimulus and the left switch to identify a left stimulus. The reverse S–R mapping was used by *S*s in Group Incompatible (16 males, 4 females); that is, *S*s pulled the right switch to identify a left stimulus and the left switch to identify a right stimulus.

Results

For analysis, each choice RT for Trials 3–400 was classified into the appropriate categories which were defined by three dichotomous variables: current prediction outcome (PO), preceding PO, and the run length of the preceding PO. Letting the

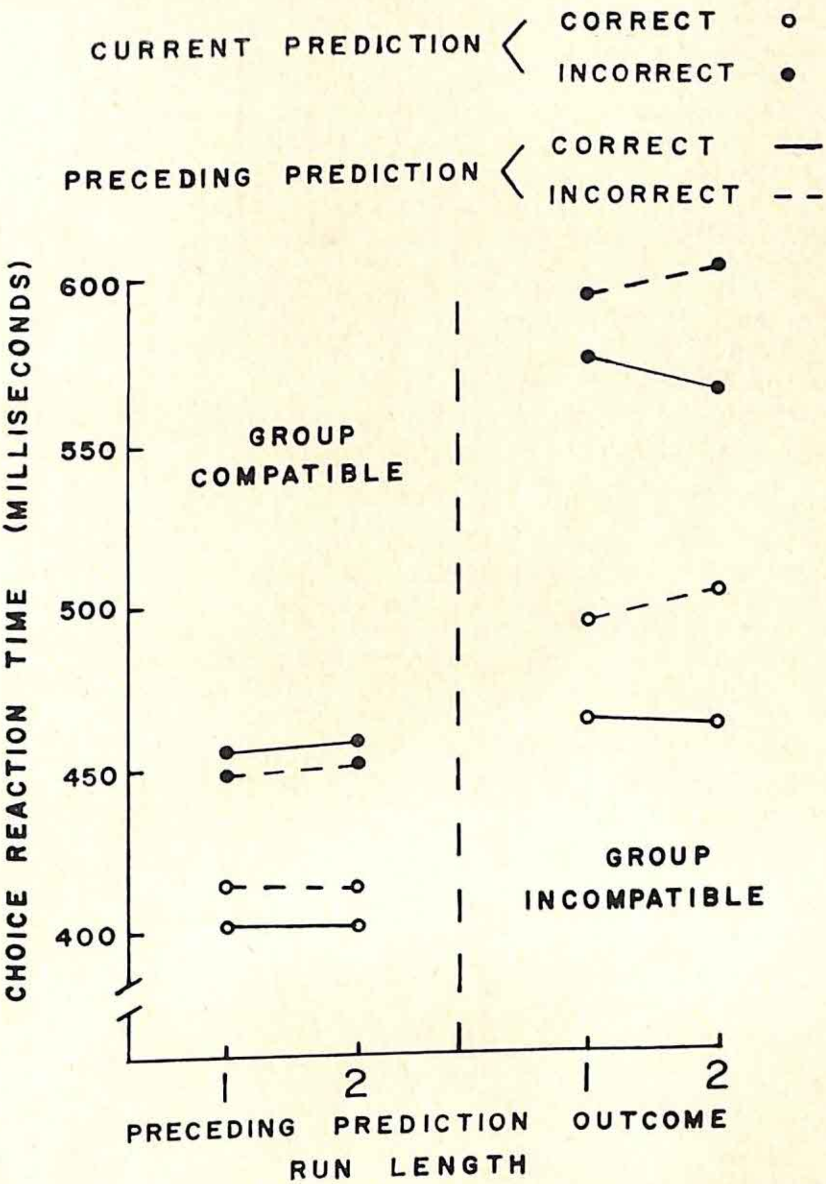

FIG. 1. Choice reaction time as a function of the prediction run length preceding predicted and nonpredicted stimuli for Group Compatible and Group Incompatible. (The prediction run length refers to the number of consecutive correct or incorrect prediction outcomes preceding the identified stimulus.)

letters "C" and "I" represent correct and incorrect predictions, the categories would be labeled: CC, CCC, IC, IIC, CI, CCI, II, III, where the last letter in the designation represents the current PO. For example, for each S the entry in Category IIC was the mean latency over all correctly predicted stimuli which were immediately preceded by two incorrect POs. After omitting error trials (i.e., anticipatory responses and incorrect identifications), each S had at least five events in each of the eight prediction run categories.

As Fig. 1 indicates, mean choice RT for Group Incompatible was longer than mean choice RT for Group Compatible. In addi-

tion, the difference between mean choice RT to correctly predicted events and mean choice RT to incorrectly predicted events was apparently greater for Group Incompatible than for Group Compatible. For correctly predicted and incorrectly predicted stimuli for each group, choice RT to a given stimulus was influenced by the PO preceding that stimulus.

The overall analysis of variance (ANOVA), a 2 (Compatibility Levels) $\times$ 2 (Current PO) $\times$ 2 (Preceding PO) $\times$ 2 (Run Length of Preceding PO) design, indicated significant main effects of compatibility, F (1, 38) = 10.32, $p < .005$; current PO, F (1, 38) = 46.24; $p < .001$; and preceding PO, F (1, 38) = 22.64, $p < .001$. The following four interactions were significant: Compatibility $\times$ Current PO, F (1, 38) = 6.89, $p < .025$; Compatibility $\times$ Preceding PO, F (1, 38) = 17.17, $p < .001$; Compatibility $\times$ Preceding PO $\times$ Run Length, F (1, 38) = 8.39, $p < .01$; and Preceding PO $\times$ Run Length, F (1, 38) = 7.20, $p < .025$. All remaining Fs were not significant at the .05 level. Consequently, the effects of PO (Current or Preceding) were reliably influenced by the level of S–R compatibility.

To examine the nature of the interactions with compatibility, a separate 2 $\times$ 2 $\times$ 2 (Current PO $\times$ Preceding PO $\times$ Run Length) ANOVA was computed for each group. For Group Compatible, a significant effect of current PO, F (1, 19) = 36.36, $p < .001$; and a significant Current PO $\times$ Preceding PO interaction, F (1, 19) = 19.85, $p < .001$ were found. None of the other Fs reached significance at the .05 level. That is, under a compatible S–R relationship a correct PO preceding a correctly predicted stimulus shortened choice RT, but a correct PO preceding an incorrectly predicted stimulus lengthened choice RT.

The results of the 2 $\times$ 2 $\times$ 2 factorial for Group Incompatible indicated significant main effects of current PO, F (1, 19) = 25.23, $p < .001$; and of preceding PO, F (1, 19) = 23.66, p < .001; but no significant main effect of run length, $F < 1$. Of the interactions, only the Preceding PO $\times$ Run Length interaction was significant, F (1, 19) = 14.31, $p < .005$. Thus, under an incompatible S–R relationship the effect of run length was dependent on the preceding PO but was not influenced by the current PO. That is, regardless of the current PO, two preceding correct POs facilitated a response to subsequent stimuli more than one preceding correct PO, and two preceding incorrect POs inhibited a response more than one preceding incorrect PO.

Discussion

The results of this experiment are consistent with the results of previously reported experiments that examined the effects of current PO on choice RT (e.g., Bernstein & Reese, 1965; Geller et al., 1971). Mean choice RT to correctly predicted stimuli was shorter than mean choice RT to incorrectly predicted stimuli. Furthermore, as observed by Keele (1969), the effect of current PO was more pronounced under an incompatible S–R relationship than under a more compatible condition. The Run Length $\times$ Preceding PO interaction for Group Incompatible replicated the finding of Whitman and Geller (1971) that the run length of the preceding PO was a significant determinant of choice RT to a particular stimulus event. On the other hand, for the Compatible condition, the run length of preceding POs did not have an effect on choice RT nor did it interact significantly with either current PO or preceding PO. Apparently the S–R relationship used by Whitman and Geller was more difficult (i.e., more incompatible) than the Compatible condition of the present experiment. Such a conclusion is reasonable since in the Whitman and Geller study the two stimulus alternatives varied on a shape dimension (i.e., "U" vs. "∩" symbols in the center of a screen) while the response set represented a location dimension (i.e., left- vs. right-hand triggers), and as implied by Smith (1968), such is an incompatible S–R relationship.

Recent investigators have suggested that expectancy theories best account for data collected under compatible S–R conditions (Hawkins et al. 1970; Hinrichs & Krainz, 1970). The correspondence of the present data of Group Compatible to the hypotheses generated by expectancy theory supported the concept of a continuous expectancy mechanism which primarily influences stimulus identification time. For Group Compatible, the Current

PO × Preceding PO interaction was significant and indicated that the mean latency in Categories CI and IC were longer than those for Categories II and CC, respectively. Thus, as hypothesized by the continuous expectancy theory, preceding correct POs increased the facilitation to identify a subsequent correctly predicted stimulus and increased the inhibition to identify an incorrectly predicted stimulus. On the other hand, incorrect POs reduced both the facilitation to identify a correctly predicted stimulus and the inhibition to identify an incorrectly predicted stimulus. However, since the Preceding PO × Run Length interaction was not significant, the hypotheses of Whitman and Geller (1971), that the effects of consecutive correct or incorrect POs should be a cumulative function of their run length, was not supported by the present findings. Thus for a compatible S–R condition stimulus identification was maximally facilitated or inhibited by two consecutive correct or incorrect POs.

Recent studies (e.g., Keele, 1969; Schvaneveldt & Chase, 1969) have demonstrated that under incompatible S–R relationships, response effects are more prominent than stimulus effects. The observed effects of preceding PO for Group Incompatible, that choice RT was a decreasing function of the number of preceding correct POs and an increasing function of the number of preceding incorrect POs regardless of the PO of the current trial, can be explained by assuming that response selection following correctly or incorrectly predicted stimuli is facilitated by preceding correct POs and inhibited by preceding incorrect POs. The significance of the Run Length × Preceding PO interaction for an Incompatible condition implied that PO had a cumulative effect over trials. That is, response facilitation or inhibition was cumulatively influenced by three consecutive correct or incorrect POs. This is in contrast with the results for Group Compatible, which indicated that only two consecutive correct or incorrect POs had an effect on choice RT to the current stimulus.

In conclusion, the results of the present experiment indicated that POs on trials preceding a given stimulus presentation in a choice RT task significantly influenced *S*s' response latency to correctly identify that stimulus and that the nature of the relationship between preceding PO and choice RT was dependent upon the mapping between the stimulus set and the response set. The differential effects of PO indicated that the level of S–R compatibility determined whether the stimulus identification component or the response-selection component of the choice reaction process was emphasized in *S*s' reaction latencies. That is, because of the ease of response selection under the compatible S–R relationship, response effects were minimized and therefore stimulus expectancy was the primary determinant of choice RT. On the other hand, an incompatible S–R relationship produced a response-selection task of such difficulty that the response effects were emphasized and the effects of stimulus expectancy on stimulus identification time were consequently masked. Given that the level of S–R compatibility determines the extent to which a reaction latency reflects the operation of a particular component of the choice reaction process, the PO × Compatibility interactions in the present study supported the hypotheses generated by a continuous expectancy theory. That is, an expectancy model is most applicable for predicting the effects of independent variables on the stimulus identification component rather than the response selection component of the choice RT process.

REFERENCES

Biederman, L., & Kaplan, R. Stimulus discriminability and S–R compatibility: Evidence for independent effects in choice reaction time. *Journal of Experimental Psychology*, 1970, **86**, 434–439.

Bernstein, I. H., & Reese, C. Behavioral hypotheses and choice reaction time. *Psychonomic Science*, 1965, **3**, 259–260.

Geller, E. S., & Pitz, G. F. Effects of prediction, probability, and run length on discrete choice reaction speed. *Journal of Experimental Psychology*, 1970, **84**, 361–367.

Geller, E. S., Whitman, C. P., Wrenn, R. F., & Shipley, W. G. Expectancy and discrete reaction time in a probability reversal design. *Journal of Experimental Psychology*, 1971, **90**, 113–119.

Hawkins, H. L., Thomas, G. B., & Drury, K. B. Perceptual versus response bias in discrete choice reaction time. *Journal of Experimental Psychology*, 1970, **84**, 514–517.

Hinrichs, J. V., & Krainz, P. L. Expectancy in choice response time: Anticipation of stimulus or response? *Journal of Experimental Psychology*, **85**, 330–334.

Keele, S. W. Repetition effect: A memory dependent process. *Journal of Experimental Psychology*, 1969, **80**, 234–248.

LaBerge, D., Legrand, R., & Hobbie, K. K. Functional identification of perceptual and re-

sponse biases in choice reaction time. *Journal of Experimental Psychology*, 1969, **79**, 295–299.

SCHAFFER, L. H. Choice reaction with variable S–R mapping. *Journal of Experimental Psychology*, 1965, **70**, 284–288.

SCHVANEVELDT, R. W., & CHASE, W. G. Sequential effects in choice reaction time. *Journal of Experimental Psychology*, 1969, **80**, 1–8.

SMITH, E. E. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, 1968, **69**, 77–110.

STERNBERG, S. Memory scanning: Mental processes revealed by reaction time experiments. *American Scientist*, 1969, **57**, 421–457.

WELFORD, A. T. The measurement of sensory motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics*, 1960, **3**, 189–229.

WHITMAN, C. P., & GELLER, E. S. Runs of correct and incorrect predictions as determinants of choice reaction time. *Psychonomic Science*, 1971, **23**, 421–423.

(Received June 26, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 305–309

STIMULUS AND RESPONSE FACTORS IN DISCRETE CHOICE REACTION TIME[1]

JAMES V. HINRICHS[2] AND JOHN L. CRAFT[3]

University of Iowa

Stimulus and response biases in choice reaction time were examined in a three-stimulus, two-response paradigm in which *S*s were required to predict stimulus presentations. Four different stimulus frequency distributions were tested in two experiments; Exp. I used a within-*S* design and Exp. II a between-*S* design. In both experiments only stimulus bias effects were found with lower frequency differences, but as stimulus frequency differences became more extreme, response biases became more predominant.

The relative contribution of stimulus identification and response execution factors to choice reaction time (RT) experiments has been studied with several methods with mixed results. Bertelson and Tisseyre (1966) manipulated response probability with stimulus probability held constant and found no response bias effect. LaBerge, Legrand, and Hobbie (1969) used a three-stimulus, two-response design, varied both stimulus and response probability, and obtained results suggesting both stimulus and response biases in choice RT. Bernstein, Schurman, and Forester (1967) also found evidence for both factors even after taking *S*s' subjective expectancies into account. Hinrichs and Krainz (1970) obtained only a stimulus expectancy effect in a three-stimulus, two-response experiment in which *S*s predicted the occurrence of the next stimulus. In that study, however, only a single set of stimulus and response probabilities was examined. One purpose of the present study was to extend the previous results to a larger set of stimulus and response probability schedules.

LaBerge et al. (1969) found an increasing response bias with an increase in the relative frequency of the most prevalent response. In a discrete trials RT experiment, Hawkins, Thomas, and Drury (1970) found evidence for response bias only with the most extreme response probability ratios (11:1), and even this difference disappeared when response accuracy was stressed. Hawkins et al. attributed the apparent response bias effects in previous studies to incorrect categorization of input stimuli. Such incorrect categorization would produce fast responses to the stimulus paired (i.e., having the same response) with the high-frequency stimulus and errors (usually eliminated from the calculation of the mean RT) to the stimulus paired with the other response.

Another factor which could lead to an apparent response bias effect in the three-stimulus design is the differential probability of expecting presentation of the two stimuli with different response probabilities. Correctly anticipated stimuli yield much faster RTs than incorrectly anticipated stimuli (e.g., Bernstein & Reese, 1965), and the expectancy effect appears to be strictly stimulus anticipation (Hinrichs & Krainz, 1970). Furthermore, at least part of the probability effect found in choice RT can be attributed to differential proportions of correct and incorrect stimulus anticipations contributing to the means producing the observed probability effect (Hinrichs, 1970; Hinrichs & Craft, 1971). Consequently, part of the response bias effect observed by LaBerge et al. (1969) might be attributable to unequal proportions of correct and incorrect stimulus anticipations.

The present experiment used the three-stimulus, two-response paradigm and also

[1] This investigation was supported by a National Institute of Mental Health Research Grant MH-16362. The authors thank Catherine Henigbaum and Robert Shaw for their assistance.

[2] Requests for reprints should be sent to James V. Hinrichs, Department of Psychology, University of Iowa, Iowa City, Iowa 52240.

[3] Now at Wartburg College, Waverly, Iowa.

required S to predict which stimulus would be presented on each trial. Denoting by A the stimulus paired with one response and by B_1 and B_2 the two stimuli paired with the other response, several probability distributions over the three stimuli ($A:B_1:B_2$) were employed ranging from 2:1:1 to 1:1:4. Unfortunately, more extreme probability differences had to be avoided to reduce the occurrence of missing cells in the data analysis. Because Ss tend to probability match their predictions to the a priori probability of presentation, infrequent stimuli are predicted infrequently, producing very few observations of the joint event.

Experiment I

Method

Subjects.—Twelve female undergraduate students served as Ss to fulfill an introductory course requirement. Ten Ss were right-handed and 2 were left-handed.

Apparatus.—The stimuli were presented on an Industrial Electronic Engineers Bina-View self-decoding readout display cell (Model KA 12/12-093-E-1886). Stimulus presentations were controlled by prepunched paper tape sequences and programmed by a solid-state electronics system. The display cell was mounted on a black partition at eye level approximately 30 cm. from the seated S. Two identical telegraph keys mounted 32 cm. below the Bina-View cell were used as response mechnisms. The response keys were operated by S's index fingers and each was identified by printed labels as corresponding to particular stimuli. Responses and latencies were recorded on punched paper tape.

Stimuli.—In each experimental session, each S received two sequences of 130 stimulus presentations. The stimuli were the digits 0, 1, and 2. During each session, one of four stimulus frequency conditions was presented. The four stimulus frequency conditions are designated by the stimulus frequency ratios 2:1:1, 1:1:1, 1:1:2, and 1:1:4, where the first number represents the relative proportion of stimuli assigned to Response A and the next numbers represent the relative proportion of the two stimuli assigned to the two B responses (designated B_1 and B_2 with B_2 the more frequent of the two in the unequal ratio conditions). Each of the 12 Ss received one stimulus frequency condition on each of 4 successive days, such that each condition was administered equally often on each day. Each S was given one of the 12 possible ways to assign three stimuli to two responses; the stimulus-response assignment for each S remained constant over the four sessions.

Procedure.—Although Ss were informed of the stimulus-response pairings, they were not told the probability of occurrence of each stimulus or the number of trials. On each trial, Ss were required to predict, as accurately as possible, the stimulus which would occur next. Upon hearing a 100-msec. warning tone, S made her prediction by saying the digit which she believed would be presented next. The prediction was recorded manually by E. The stimulus digit was then presented 2 sec. after the onset of the warning tone. The Ss were instructed to respond to the stimulus by pressing the corresponding key as fast as possible while maintaining accuracy. The response terminated the presentation of the stimulus and then initiated the start of the next trial after a 2-sec. delay. The S was required to make her prediction of the presentation of the next stimulus after the onset of the warning tone and before the stimulus was presented. The first 10 trials in the 130-trial sequence were considered warm-up trials and were not analyzed. Over the last 120 trials in each sequence the appropriate stimulus ratios were maintained.

Results and Discussion

In each frequency condition each S contributed 240 observations distributed according to the presentation probability and S's prediction probability. The mean RTs in each experimental condition and the relative proportion of responses in each cell are presented in Table 1. The presented RTs are the means of the 12 individual S median RTs in each condition. The proportion of errors for each frequency condition and stimulus are also shown in Table 1.

The RTs were analyzed by a $3 \times 3 \times 4$ analysis of variance with presented stimulus, predicted stimulus, and frequency condition as factors. The main effect of frequency condition was not significant, $F < 1$, nor was the Frequency $\times$ Predicted Stimulus interaction, $F < 1$, but the Frequency $\times$ Presented Stimulus interaction was significant, $F\ (6, 66) = 7.45$, $p < .001$. Significant contributions to the overall variance were also produced by the factors of presented stimulus, $F\ (2, 22) = 24.27$, $p < .001$; predicted stimulus, $F\ (2, 22) = 10.93$, $p < .01$; and their interaction, $F\ (4, 44) = 57.06$, $p < .001$.

Inspection of the pattern of RTs in Table 1 reveals the nature of the various main effects and their interactions. As in

TABLE 1

Mean Reaction Times and Standard Errors (in msec.), Proportion of Observations, and Error Rate as a Function of Stimulus Ratio, Stimulus Presented, and Stimulus Predicted in Experiment I

Performance measure	Stimulus predicted	2:1:1 A	2:1:1 B_1	2:1:1 B_2	1:1:1 A	1:1:1 B_1	1:1:1 B_2	1:1:2 A	1:1:2 B_1	1:1:2 B_2	1:1:4 A	1:1:4 B_1	1:1:4 B_2
Reaction times	A	421 (9)	507 (15)	496 (16)	425 (25)	503 (28)	506 (30)	420 (11)	502 (14)	494 (14)	452 (25)	527 (43)	487 (29)
	B_1	525 (18)	419 (13)	522 (23)	545 (29)	421 (25)	507 (27)	575 (23)	402 (11)	491 (14)	626 (46)	442 (42)	482 (33)
	B_2	530 (15)	532 (22)	408 (9)	560 (28)	514 (27)	418 (30)	563 (21)	517 (14)	408 (11)	603 (38)	540 (42)	400 (23)
	Overall median	489 (15)	487 (16)	473 (15)	512 (27)	479 (25)	473 (27)	532 (20)	480 (15)	458 (11)	578 (39)	515 (41)	429 (25)
Proportion of observations	A	.188	.092	.099	.110	.106	.096	.061	.062	.140	.037	.037	.129
	B_1	.134	.080	.064	.101	.098	.109	.065	.079	.151	.026	.041	.166
	B_2	.160	.070	.082	.103	.126	.123	.102	.107	.202	.078	.085	.365
	Sum	.482	.242	.245	.314	.330	.328	.228	.248	.493	.141	.163	.660
	Errors	.018	.008	.005	.019	.003	.005	.022	.002	.007	.026	.004	.007

Note.—Standard errors in parentheses.

the Hinrichs and Krainz (1970) study, there is a clear separation between RTs to correctly and incorrectly predicted stimuli in all conditions. Furthermore, consistent with the Hinrichs and Krainz results, in the 2:1:1 and 1:1:1 conditions there is little difference among RTs to stimuli correctly predicted and among RTs to incorrectly predicted stimuli, regardless of the particular stimulus predicted or presented. However, as the discrepancy between the presentation frequency increases, as in Cond. 1:1:2 and 1:1:4, the nature of the prediction and the stimulus presentation influences the observed RT. In the 1:1:2 condition (and to a lesser extent even in the 1:1:1 condition), RTs are slower to the incorrect prediction of the single stimulus (associated with Response A) than to the incorrect prediction of either of the other two, more frequently presented, stimuli (associated with Response B). In the 1:1:4 condition this effect is even more marked, so that RTs to the less frequent of the paired stimuli (B_1) are slower than to the more frequent stimulus (B_2).

The systematic change in RTs with increasing differences in the frequency of B_1 and B_2 stimuli suggests the presence of a response bias effect. Evidence for a response bias is especially clear when *S*'s predictions are disregarded as in the overall medians presented in Table 1. Although the presented frequency differences are not so large, the pattern of results is very similar to those reported by LaBerge, Legrand, and Hobbie (1969) across stimulus frequency variation (cf. data in Table 1 with Fig. 1 in LaBerge et al.).

Both the present results and the LaBerge et al. (1969) results are potentially contaminated by the use of a within-*S* design. Because all *S*s participate in all conditions, there exists the possibility of "carry over" of biases or influences from other conditions in the experiment. Therefore, it was considered prudent to replicate the present study using a between-*S* design.

Experiment II

Method

The details of the second experiment were the same as in the first experiment except that 12 different *S*s were employed in each condition. Three *S*s, 2 in the 1:1:4 condition and 1 in the 1:1:2 condition, produced data with no observations in some cells; to allow statistical analysis they were replaced.

TABLE 2

Mean Reaction Times and Standard Errors (in msec.), Proportion of Observations, and Error Rate as a Function of Stimulus Ratio, Stimulus Presented, and Stimulus Predicted in Experiment II

Performance measure	Stimulus predicted	Experimental cond. and stimulus presented											
		2:1:1			1:1:1			1:1:2			1:1:4		
		A	B_1	B_2	A	B_1	B_2	A	B_1	B_2	A	B_1	B_2
Reaction times	A	491	583	584	464	553	548	507	578	567	471	578	519
		(30)	(36)	(32)	(18)	(13)	(16)	(27)	(30)	(25)	(15)	(13)	(18)
	B_1	586	495	607	589	471	557	635	501	565	640	482	542
		(32)	(27)	(35)	(15)	(16)	(20)	(33)	(28)	(34)	(21)	(22)	(18)
	B_2	576	616	493	595	555	441	624	573	485	599	603	436
		(29)	(36)	(29)	(14)	(17)	(15)	(23)	(36)	(31)	(17)	(21)	(17)
	Overall median	539	565	560	551	526	511	596	554	530	583	573	464
		(30)	(34)	(29)	(14)	(11)	(15)	(29)	(32)	(33)	(15)	(13)	(17)
Proportion of observations	A	.217	.119	.122	.108	.114	.108	.061	.061	.132	.034	.031	.125
	B_1	.130	.058	.049	.099	.104	.104	.063	.068	.154	.030	.027	.103
	B_2	.140	.066	.071	.100	.107	.106	.104	.116	.203	.093	.107	.434
	Sum	.487	.243	.232	.307	.325	.318	.228	.245	.489	.157	.165	.662
	Errors	.013	.007	.018	.026	.008	.015	.022	.005	.011	.010	.002	.005

Note.—Standard errors in parentheses.

Results and Discussion

The data were analyzed in the same manner as in Exp. I and the comparable items of interest are presented in Table 2. A 3 × 3 × 4 × 2 analysis of variance was used to analyze the data; the additional factor not included in the analysis of Exp. I was the effect of the two stimulus lists. The effect of stimulus list was a significant source of variance, F (1, 44) = 21.78, $p < .001$, indicating a significant reduction in mean RTs in the second list, but it did not interact significantly with any other factor. Otherwise, the statistical results of Exp. II exactly paralleled the results of Exp. I. The main effects of presented and predicted stimuli were significant, F (2, 88) = 31.88, $p < .001$, and F (2, 88) = 14.26, $p < .001$, and their interaction, F (4, 176) = 141.23, $p < .001$; as was the Frequency × Presented Stimulus interaction, F (6, 88) = 8.65, $p < .001$. Furthermore, comparison of Table 1 and Table 2 reveals generally similar patterns of results.

General Discussion

In general, the present results do not contradict any of the previous results cited earlier. That is, with lower frequency differences among the stimuli, only stimulus bias effects are found. Consistent with the previous application of the verbal prediction procedure in the three-stimulus, two-response paradigm, there is little or no advantage in anticipating a stimulus with the same response as the stimulus presented (Hinrichs & Krainz, 1970). However, when response probabilities became more discrepant, a response bias did occur in the present study in the same sense as in the LaBerge, Legrand, and Hobbie (1969) study. That is, averaging over S's predictions, the responses to the stimulus (B_1) with the same response as the high-probability stimulus (B_2) were executed faster than to the stimulus (A) with the same probability as B_1 but associated with a different response. Nevertheless, even in the most discrepant frequency condition (1:1:4) when a response bias in the LaBerge et al. (1969) sense is clearly evident, there is still no prediction advantage for stimuli sharing the same response.

Hence, something of a paradox exists: by one reasonable index (differences in mean RT across frequency levels), evidence is found for both stimulus and response biases in choice RT; but by another criterion (depending on separation by verbal anticipations of stimuli), only a stimulus bias is found. One interpreta-

tion which has been suggested (Hawkins et al., 1970) attributes the response bias factor to inclusion of (unobservable) erroneous fast responses to the stimulus associated with the high-probability response. Although this effect may play a role, it would also appear to predict that anticipation of the stimulus associated with the presented stimulus (e.g., predicting B_2 when B_1 is presented) would produce faster RTs than predicting the stimulus associated with the other response (e.g., predicting A when B_1 is presented). This prediction is not supported by the present data.

The obtained results for correctly anticipated stimuli are generally consistent with a stimulus bias effect, with little response bias even in the most extreme frequency condition. In other words, the B_1 RTs are closer in value to the A RTs than to the B_2 RTs. However, the response bias is most obvious in the incorrectly anticipated cases where the B_1 responses, although consistently slower than the B_2 responses are closer to the B_2 RTs than to the A RTs. Therefore, it may be the case as noted in other studies (e.g., Bernstein & Reese, 1965; Hinrichs & Craft, 1971), that probability effects and response biases occur only for incorrectly anticipated stimuli.

REFERENCES

BERNSTEIN, I. H., & REESE, C. Behavioral hypotheses and choice reaction time. *Psychonomic Science*, 1965, **3**, 259–260.

BERNSTEIN, I. H., SCHURMAN, D. L., & FORESTER, G. Choice reaction time as a function of stimulus uncertainty, response uncertainty, and behavioral hypotheses. *Journal of Experimental Psychology*, 1967, **74**, 517–524.

BERTELSON, P., & TISSEYRE, F. Choice reaction time as a function of stimulus versus response relative frequency of occurrence. *Nature*, 1966, **212**, 1069–1070.

HAWKINS, H. L., THOMAS, G. B., & DRURY, K. B. Perceptual versus response bias in discrete choice reaction time. *Journal of Experimental Psychology*, 1970, **84**, 514–517.

HINRICHS, J. V. Probability and expectancy in two-choice reaction time. *Psychonomic Science*, 1970, **21**, 227–228.

HINRICHS, J. V., & CRAFT, J. L. Verbal expectancy and probability in two-choice reaction time. *Journal of Experimental Psychology*, 1971, **88**, 367–371.

HINRICHS, J. V., & KRAINZ, P. L. Expectancy in choice reaction time: Stimulus or response anticipation? *Journal of Experimental Psychology*, 1970, **85**, 330–334.

LABERGE, D., LEGRAND, R., & HOBBIE, R. K. Functional identification of perceptual and response biases in choice reaction time. *Journal of Experimental Psychology*, 1969, **79**, 295–299.

(Received June 29, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 310–317

PREDICTION PROBABILITY AS A DETERMINER OF ANTICIPATORY AND PREPARATORY ELECTRODERMAL BEHAVIOR [1]

WILLIAM W. GRINGS [2] AND HARRIET I. SUKONECK

University of Southern California

Four electrodermal responses during discrimination conditioning were observed as a function of the probability that the CS would be followed by the UCS (a shock). The responses were those immediately following CS, just prior to UCS onset, following UCS onset (UCRs), or occurring with omission of the UCS. Probabilities of UCS following CS were 0%, 25%, 75%, and 100%. The responses just prior to UCS onset, termed anticipatory responses (or ARs), and the UCS omission responses increased in magnitude as probability of shock increased. UCR magnitudes decreased as probability increased. Parallel results were obtained when a measure of subjective "certainty" was employed.

In paired stimulation situations, like classical conditioning and reaction time tasks with a warning signal, characteristic autonomic reactions are elicited by the first stimulus and determine responses to the second stimulus. Such responses are likely to be labeled conditioned responses (CRs) in the conditioning context, and they are assumed to prepare the organism for the receipt of the second stimulus in the reaction time task. Experimental research employing the GSR has examined the topography of the response in such situations and has identified a number of response components which become evident when the interstimulus interval (ISI) exceeds the shortest latencies of the response.

With ISIs of 5 sec. or longer, four variables are commonly identified. The first is a response following closely the onset of the signal stimulus. It probably reflects a combination of orienting and anticipatory behavior. The second response occurs in a restricted time interval just prior to receipt of the second stimulus and is assumed to reflect anticipatory preparation for receipt of that stimulus. The third is the response to the second stimulus. It is influenced by the preparatory state of the individual and is predicted to be smaller in magnitude when the organism is prepared than when he is not prepared. The fourth response is measured at the point in time where the response to the second stimulus would occur but is observed in the absence of the second stimulus, (e.g., Grings, Lockhart, & Dameron, 1962).

Various terminologies have been employed to designate the four responses. One of the first investigations to compare the four (Stewart, Winokur, Stern, Guze, Pfeiffer, & Hornung, 1959) termed the first an orienting response (OR), the second an anticipatory response (AR), the third the unconditioned response (UCR), and the fourth the response to UCS during extinction (i.e., the response to the omitted UCS). Another terminology, recommended by Lockhart (1966) identifies the responses in the order named above as (*a*) CS response, (*b*) pre-UCS response, (*c*) UCR, and (*d*) UCS-omission response. In the material which follows both of the above terminologies will be used, with the designations being OR, AR, UCR, and UCS-omission response.

The present research assumes that the relative magnitudes of the different responses will reflect the information conveyed by the first stimulus (signal or CS) about the second stimulus (usually noxious) to follow. Differences in ability to predict the occurrence and properties of the second stimulus, as well as the actual properties

[1] This research was supported in part by Grant MH 03916 from the National Institute of Mental Health.

[2] Requests for reprints should be sent to William W. Grings, Department of Psychology, University of Southern California, University Park, Los Angeles, California 90007.

of that stimulus, are assumed to produce characteristic patterns of response. In order to evaluate the above assumptions, the present experiment manipulated prediction probability, or the likelihood that the second stimulus would follow the first stimulus. Four distinctive first stimuli were each followed by a different probability of appearance of a second (electric shock) stimulus, and the various electrodermal responses were measured.

One definition of prediction probability is in terms of the physical proportion of training trials on which the particular cue stimulus is followed by the shock. Four categories of such physical probability are used. The second definition is in terms of subjective statements by *S* of his expectation about receipt of shock on a particular trial. A third definition refers to the subjective judgment of confidence or belief *S* has in his statement of subjective probability. The general prediction is that the first two electrodermal responses (the CS response, or OR; and the pre-UCS response, or AR) will increase in amplitude with training and with probability, as will also the UCS-omission response (which is assumed to be a CR). The third response (or UCR) is predicted to decrease in magnitude with training and with probability.

METHOD

Subjects.—The *S*s were 32 male undergraduate students who participated in the experiment for credit in their introductory psychology course. Primary comparisons were made within *S*s.

Apparatus and materials.—There were four CSs produced by combinations of two dimensions, color (red or blue) and shape (horizontally arranged dots or vertically arranged dots), projected by a Grason-Stadler multistimulus projector (Model E4580-3) resting on a table in front of *S*. The UCS was a DC shock from a Grass Model 5 stimulator administered to the volar surface of the right arm through 1.7-cm. circular electrodes (spaced 2.5 cm. from center to center). The CS duration of 10 sec., CS–UCS interval of 10 sec., and UCS duration of .5 sec. were controlled by Hunter timers. A Gerbrands programmer controlled the intertrial intervals systematically varied among 30, 35, 40, and 45 sec. The *E* controlled manually whether or not a shock occurred on a particular trial.

The shock (UCS) schedules (proportion of shock trials) for the various CSs were .00, .25, .75, and 1.00. Assignment of the shape-color properties to the probabilities was rotated through the use of four subgroups of *S*s. Four different trial orders of stimulus-probability presentations were used. Order was random in blocks of 16 such that for each 16 trials each of the 4 stimuli appeared four times in random order. There were 80 training trials in all. After the forty-eighth trial, half of the *S*s were given explicit instructions about the stimulus probability relationships to facilitate accurate discrimination. The other *S*s were not instructed but had their electrodes checked by *E* at this point.

The GSR, recorded on a Beckman Type R Dynograph, was obtained as a DC resistance change through 1.27×1.59 cm. silver electrodes coated with electrode paste and taped to the first and third fingers of *S*'s left hand. The input circuit was a modified (Darrow) Wheatstone bridge with current of $45 \mu a$.

The *S*s indicated their guesses of probability of shock by pressing one of five levers, labeled 0, 25, 50, 75, and 100, which were mounted in a small box on the table. Another small box contained four buttons labeled ABSOLUTELY UNSURE, RELATIVELY UNSURE, RELATIVELY SURE, and ABSOLUTELY SURE, which *S* operated to indicate how certain he felt of his prediction about the probability of shock.

Procedure.—Each *S* was told that he would be shown various combinations of color and form followed sometimes by shock but not always, and that his task would be "to learn to report the probability that a stimulus signals shock to follow." The GSR was explained to him, electrodes were attached, and he was shown how to indicate his probability estimate by pushing one of the five levers and his confidence that his guess was correct by pushing one of the four other buttons. He was told to push the button immediately after pushing one of the levers and to make his lever response as soon as the light stimulus was presented.

After these preliminary instructions a shock work-up procedure was carried out to determine the intensity of UCS for each *S*. Then the elements of the CSs (two shapes and two colors) were presented once in random order before *E* returned to the adjacent room and began the training trials. At the conclusion of the 80 training trials, a written questionnaire was completed by *S* in conjunction with receipt of one presentation of the four CSs. This provided a final indication of *S*'s guess about the probability of shock following each CS, a rating of how uncomfortable the shock was for each CS, and a statement of how many times *S* believed he was shocked for each CS.

RESULTS

Four electrodermal magnitudes were determined: the maximum change between 1 and 3.5 sec. after signal onset; the maximum change between 3.5 and 10.5 sec. after signal onset; the maximum change

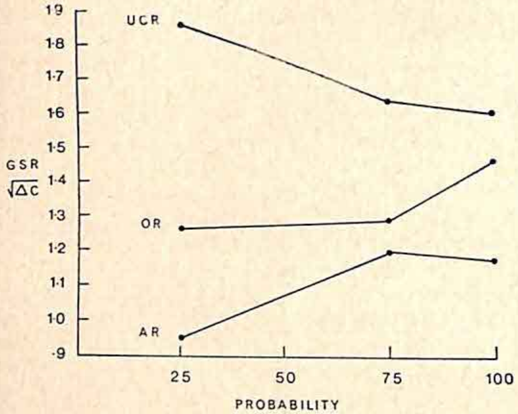

Fig. 1. Magnitude of various responses as a function of stimulus probabilities.

between 1 and 6 sec. after shock (UCS) onset; and the maximum change between 1 and 6 sec. after the point of UCS onset, but occurring on trials where shock was omitted. For convenience the responses will be labeled OR, AR, UCR, and UCS-omission response, in the order named. All magnitudes are expressed as the square root of the change in conductance in micromhos.

Since UCRs are available only on trials where shock is administered and UCS-omission responses occur only when shock is omitted, certain preliminary decisions were necessary in selecting comparison trials. For estimation of magnitude trends in ORs, ARs, and UCRs, a representative trial for each CS in each block of 16 trials was chosen as follows. The trial used for the 25% stimulus was the one out of 4 on which shock was presented. Then for the 75% and 100% stimuli, the corresponding trial in terms of trial number was used. It was found that this procedure did not yield ORs or ARs significantly different from those based on the average of four responses to each stimulus for each trial block of 16 trials. A similar procedure was used for evaluating the UCS-omission response. In this case the criterion trial was determined by the 1 nonshocked trial in the 75% group. For comparisons involving UCRs the 0% probability stimulus was omitted, and for comparisons involving UCS-omission responses the 100% stimulus was omitted.

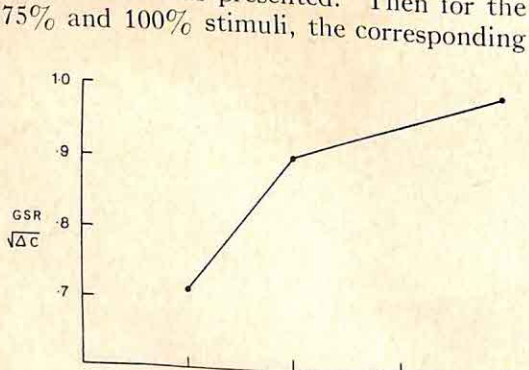

Fig. 2. The UCS-omission response magnitude for different stimulus probabilities.

A Trials × Probability analysis of variance was done, for each of the responses, over the first four trial blocks (the period before instruction variation was introduced). For the OR, only the probability effect was significant, F (2, 341) = 5.40, $p < .01$, with responses increasing in magnitude as probability increased. For the AR, the probability effect was also significant, F (2, 341) = 7.70, with responses increasing in magnitude as a function of increasing probability. For the UCR, responses decreased significantly as probability increased, F (2, 341) = 9.67, and the Probability × Trials interaction was also significant, F (6, 341) = 3.67, $p < .01$. With the 25% and 75% stimuli, the UCR increased in magnitude over trials and with the 100% stimulus it decreased. The overall effect of probability upon the first three responses (OR, AR, and UCR) is shown in Fig. 1. The effect of probability on the UCS-omission rseponse was also significant, F (2, 341) = 6.23, $p < .01$. These data are shown in Fig. 2. To evaluate the effect of instructions introduced to half of the *S*s after the fourth trial block, a comparison across Trial Blocks 4 and 5 was made for each response. For the OR, there were no significant differences due to instructions or probability. For the AR, the effect of instructions was significant, F (1, 30) = 5.37, $p < .05$, with the instructed group having higher response magnitudes than the uninstructed group. The overall probability effect for the AR was significant F (2, 60) = 3.41, $p < .05$, with response magnitudes increasing as a function of increasing probabilities; and there

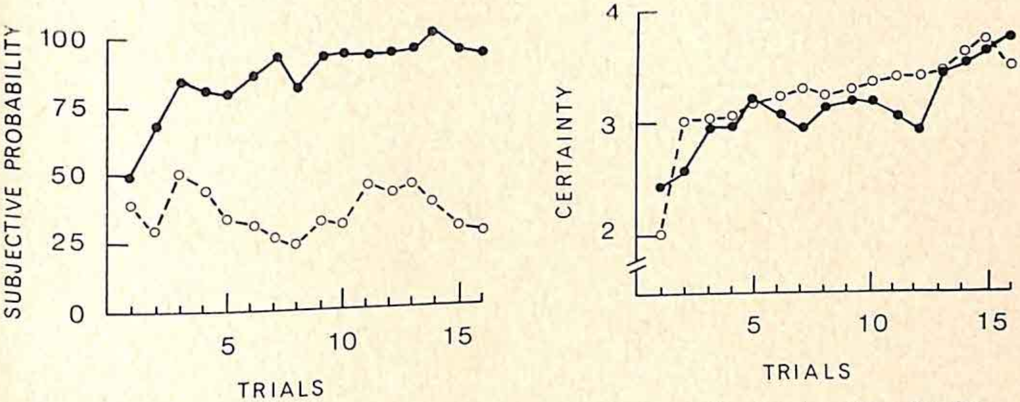

FIG. 3. Growth over trials of average subjective probability and certainty ratings for the 25% (broken lines, ---) and the 100% (solid lines, ——) stimuli.

was a significant overall increase in AR magnitude between trial blocks, F (1, 30) = 13.69. There was also a significant Instruction × Trials interaction for the AR, F (1, 30) = 5.24, $p < .05$. For the UCR, only the probability effect was significant, F (2, 60) = 12.41, with response magnitude decreasing as probability increased. The UCR-omission response yielded a significant probability effect, F (2, 60) = 6.69, $p < .01$, and an Instruction × Trials interaction, F (1,30) = 4.43, $p < .05$.

Subjective probability and certainty variables are difficult to assess for several reasons. It is not possible to separate completely subjective from objective probability, for *S*s learned very rapidly to predict the actual probability associated with each stimulus. As can be seen from Fig. 3, the 25% and 100% probabilities were learned in the first trial block. The certainty judgment value was also closely related to trials. For example, the mean certainty value over all probabilities on Trial 1 was 2.1 (2 standing for RELATIVELY UNSURE). By the fourth presentation of each stimulus this mean certainty value over all probabilities had advanced to 3.2 (with 3 indicating RELATIVELY SURE). By the sixteenth presentation of each stimulus the mean had become 3.7. The course of development of certainty for the 25% and 100% stimuli can also be seen in Fig. 3.

To test the prediction that magnitude of response to the shock would decrease with greater certainty of subjective judgment within a given objective probability class, *S*s were segregated into three "certainty" groups, formed on the basis of prediction responses during the last four trials on the 100% stimulus prior to instruction manipulation. The first group, labeled CERTAIN, met the criterion of giving both correct subjective probability estimates and ratings of ABSOLUTELY SURE on all four trials. The other extreme, or UNCERTAIN group, consisted of *S*s who on all four trials displayed either an incorrect probability prediction or a certainty value below that of ABSOLUTELY SURE. The remaining *S*s were in an intermediate group.

Using the two extreme groups of CERTAIN and UNCERTAIN *S*s, the mean magnitude of response to the shock was computed for the four trials. For the CERTAIN group the mean was 1.32 micromhos; and for the UNCERTAIN group the mean was 1.98 micromhos. When this difference was evaluated by a Mann-Whitney U-test, the difference was significant at the .01 level. This comparison suggests that the ability to predict with assuredness the occurrence of the shock (within a given physical probability category) is related to the magnitude of response elicited by the shock.

DISCUSSION

Given the facts of an increase in AR magnitudes and a decrease in UCR magnitude with increases in prediction probability, it becomes necessary to explain the basis for the relations. Probably the simplest explanation

is that higher probability of shock leads to the learning of a higher expectation of being shocked, the higher expectation of shock leads to various forms of anticipatory behavior (e.g., larger ARs), and this somehow produces a lower response to the UCS.

The above interpretation raises the question of how the AR relates to (or perhaps determines) the UCR magnitude. Two classes of explanation exist. One is that the AR represents the crucial response which is sensitive to the independent variation and that the reduced magnitude of the UCR is a biproduct of the AR. There is some experimental evidence, for example, to show that in a nonlearning paired-stimulation situation, the response to a standard second stimulus is a function of the amplitude of response to the first stimulus at the time of onset of the second stimulus (Grings & Schell, 1969).

The second type of explanation is that both the increased AR and the decreased UCR represent some common process, i.e., some form of general learned behavior. The authors have favored the interpretation that the AR and UCR changes represent a common process of preparation for the receipt of the second stimulus (Grings, 1969). Others have suggested that the process is one of conditioned inhibition (Kimble & Ost, 1961; Kimmel, 1966). One further interpretation (Lykken, 1962, 1968) assumed the operation of an inhibitory–facilitory mechanism set up by the CS in a process termed "preception."

The differences among the various explanations will not be reviewed here. Instead, the assumptions stated earlier in this article will be reviewed and elaborated in light of the data, as follows. Two stimuli are assumed to differ in significance to the organism (at the simplest level they differ in intensity). When temporal arrangements are optimal (weak stimulus, or CS, preceding the more significant stimulus, or UCS), the weak stimulus becomes a signal in the sense of informing the organism of the approaching occurrence of the UCS. The CS comes to elicit an interpretive or perceptual reaction, one feature of which is to produce responses of anticipation or preparation for receipt of the UCS. When the UCS is noxious, it is reasonable to assume that the preparatory reaction serves to control or modulate the arousal effects of the UCS.

Placed in the above context, the OR is seen as a response that is at first a reflection of *S* alerting to stimuli, with differences attributable to intrinsic significance of the stimuli. If the CS is followed immediately by the other stimulus (i.e., the UCS), the OR habituates more slowly with repeated presentation than if it retains less or no signal significance. (This form of differential habituation could be manifested as differential OR responding to a CS + and a CS −, a phenomenon designated by some as conditioned ORs). The UCS in the same context is seen as strong enough to elicit a defensive reaction. The terms orienting reaction and defensive reaction are used basically in the same sense as proposed by Sokolov (1963).

Further, it is assumed that the AR is an anticipatory and adaptive response resulting from *S*'s perception of the signal characteristic (meaning or information) of the CS. It is assumed that an AR preceding a strong stimulus may effectively reduce the disequilibrium produced by that stimulus (although under some circumstances it might have an opposite effect). In a general discussion of these problems some time ago, Berlyne (1960, p. 186) suggested that a signal stimulus which precedes an event with high arousal value has the following beneficial features:

> (1) Arousal increases before the heralded event makes its impact, thus preparing the organism for prompt action and optimal receptivity for information. (2) Arousal increases gradually rather than sharply, thus minimizing the disturbance of internal equilibrium. (3) Arousal does not rise unnecessarily high once the anticipated event has appeared, thus minimizing wear and tear and upheaval. (4) Arousal falls to a normal or optimal level as soon as the rise in arousal has performed its function.

The emphasis in all of the above is upon the information which the CS provides about the UCS, and upon the perceptual reaction elicited by the CS. Major classes of information being processed by *S* include the properties of the UCS and its likelihood (or certainty) of occurrence. The latter variable, which was manipulated in this experiment, had been designated as "event" certainty to differentiate it from other forms, like "time" certainty (based on predictability by *S* of time of occurrence of the UCS) and "quality" certainty (based on predictability of stimulus quality and intensity). Such prediction variables have been shown to be important in related contexts where strong stimuli, like electric shocks, are administered with warning signals and autonomic responses are observed (e.g., Badia, McBane, Suter, & Lewis, 1966; Epstein & Clark, 1970). The central conclusion from the present data is that prediction or infor-

mational certainty variation is the underlying determiner for the results observed in Fig. 1 and 2.

Other considerations arise from the fact that the GSR wave in such situations is actually a complex in which different response components can be identified. An early classification (Grings et al., 1962) used six different response "forms" when a 5-sec. CS–UCS interval was used. Attention to response topography has led to tendencies to ascribe behavior meaning to the various "humps" of the GSR wave (as has been done tentatively in the present paper). That leads to study of interrelations of magnitude, amplitude, and frequency aspects of the response components and the assertion that certain measures more effectively represent the conditioned behavior than do others (e.g., Prokasy & Ebel, 1967). The evidence to date favors an interpretation that the AR is the most sensitive to conditioning variation; yet numerous studies also report conditioning of the OR. The present study was not designed to test these differences among component responses but rather to demonstrate a change in total form of responding which could be assumed to reflect learned behavior.

On the other hand, the research was designed to eliminate nonlearning changes attributable to the independent variation, such as might occur with variation of CS intensity or interstimulus interval (e.g., Grings & Schell, 1969). It was recognized that an independent variable which had the intrinsic capability of varying magnitude of the OR or AR might alter the response complex in ways not interpretable as learned expectancy or preparatory behavior. To prevent this, conditioned stimuli were chosen which had been previously demonstrated to elicit comparable orienting GSRs, and as further precaution the stimuli were systematically varied in assignment to probability classes from *S* to *S*. It was also assumed that there is nothing in the probability manipulation per se (i.e., independent of a learning process) that would lead to differences in response to the signal stimulus at the outset. That assumption seems to be justified.

It may still be of interest to determine which of the various response elements contributes the most information, or how one might define a complex response pattern to best reflect the learning that is occurring. That is a difficult task which extends beyond the capability of the present study. Nevertheless, some exploratory analyses were undertaken to evaluate the degree of relation (or independence) among the three variables (OR, AR, and UCR).

A first approach employed intercorrelations among the component responses. If the reduction in magnitude of the UCR with higher levels of probability occurs only because of variation in the AR (and if there is some kind of total effector response capability), there should be a negative correlation between those two components. Such a correlation (– .54) has been shown to exist where the first of the two successive responses was purposively manipulated in a nonconditioning situation. On the other hand, when factors producing such peripheral response interference were minimized, the correlation was positive (+ .57) (Grings & Schell, 1969). The high positive correlation was assumed to be mediated by an underlying responsivity variable which would contribute to an interindividual correlation.

In the present study several correlations between UCRs and ARs were computed (single trials, averages of trials, within probability classes and across such classes, during early trial periods, and during later periods). The general results were low positive values. The largest positive value (.48) was obtained when optimum circumstances for maximizing reliability were present (average of last four trials in the 100% probability situation). Correlations of ORs and UCRs centered about zero. Reliability estimates obtained from a correlation of UCRs ranged from .59 for adjacent single trials early in the learning period to .73 for averages of two adjacent trials during the fourth block of the 100% probability.

Because of the small number of cases and trials for estimating the above correlations, similar computations were made on a related set of data (Grings & Schell, 1971) where each *S* had uniform stimulational experience for 30 CS–UCS trials. The correlation between AR and UCR (employing average responses over all trials) was significantly positive, $r\ (58) = .33,\ p < .01$. The correlation between OR and UCR magnitudes was also significantly positive (.53) as was the correlation between OR and AR (.62). The partial correlation of AR and UCR with OR held constant was insignificant (– .003). Assuming the OR is a measure of general responsivity, one may conclude that the correlation between AR and UCR is largely a result of differences in individual reactivity.

Since the present experiment was run, additional relevant data have become available.

Slubicka (1971) averaged OR, AR, and UCR amplitudes over a series of 16 reinforced trials, obtaining correlations of + .45 between ARs and UCRs; + .57 between ORs and UCRs; and + .55 between ORs and ARs. The partial correlation between AR and UCR amplitudes with OR amplitude held constant is + .19 (not significant for $N = 40$). Of special interest in Slubicka's results is the inclusion of frequency and latency variables. The highest correlations obtained were between AR frequency and AR onset latency ($r = +.83$) and between AR frequency and AR latency of peak response ($r = +.79$), suggesting that the more that ARs occur the later are their onsets and peaks. Further, the only significant negative correlations obtained were between UCR frequency and AR amplitudes ($r = -.73$) and between UCR frequency and AR peak latency ($r = -.52$). Thus the presence of late ARs may be seen to reduce the frequency of UCRs. The largest correlations with UCR amplitude was the previously mentioned value of + .57 with OR amplitude, suggesting the presence of an underlying individual responsivity factor.

Intercorrelations like the above illustrate again the fact that many variables are involved in the definition of a conditioned electrodermal response, particularly the existence of an individual difference reactivity level factor, the relevance of latency criteria in defining frequency measures, and the possible differences between amplitude and magnitude measurement. Approaches other than those mentioned above probably merit scrutiny. For example, the method of multiple correlation and stepwise regression may be helpful.

Other evaluations from the present study may be made from subsamples of data.[3] When faced with the above problems of estimating UCR magnitudes which are independent of anticipatory responses in a situation similar to the present one (eyeblink conditioning), Kimble and Ost (1961) based their conclusions about UCR changes on trials where an AR did not occur. Even though that procedure presented methodological difficulties (which the authors discuss), Kimble and Ost judged it to be adequate, and UCR diminution was observed to occur on the selected trials. In the present study, such a comparison is unfeasible because of the high frequency of ARs. For example, on the most relevant trial block (the acquisition block just before instruction variation was introduced) only 5 of the 32 *S*s failed to give ARs on the 25% trial which received a UCS. Of these, only 2 failed to give an AR on the corresponding 100% trial. It thus becomes impractical to examine the UCR changes as a function of probability in the absence of ARs. The closest one can come to estimating the effect of AR on UCR by *S* selection is to restrict observation to the 100% stimulus (where there are four trials per block), then to use the last two acquisition blocks and locate a sample of *S*s having at least one trial without an AR. There were 14 such *S*s. When the UCRs on the trial without the AR are compared with the UCRs on the adjacent trial having an AR the mean difference is .072, and $t\ (13) = .55$.

If the 100% UCRs and the 25% UCRs for these 14 *S*s are examined, the difference between the mean values over *S*s, while in the same direction as for the total sample, is not significant, mean difference = .168, $t\ (13) = 1.68$. However, the mean difference compares well with the mean difference over all *S*s (.220), and the error term is such that if the number of *S*s in the subsample had been as large as that in the total group, the value of .168 would have been significant. Thus, it seems reasonable to conclude that these 14 *S*s are not systematically different from the other 18 *S*s with respect to the effect of shock probability on UCR magnitude.

[3] The authors are grateful to Anne M. Schell for assistance with this analysis.

REFERENCES

Badia, P., McBane, B., Suter, S., & Lewis, P. Preference behavior in an immediate versus variably delayed shock situation with and without a warning signal. *Journal of Experimental Psychology*, 1966, **72**, 847–852.

Berlyne, D. E., *Conflict, arousal and curiosity*. New York: McGraw-Hill, 1960.

Epstein, S., & Clarke, S. Heart rate and skin conductance during experimentally induced anxiety: Effects of anticipated intensity of noxious stimulation and experience. *Journal of Experimental Psychology*, 1970, **84**, 105–112.

Grings, W. W. Anticipatory and preparatory electrodermal behavior in paired stimulation situations. *Psychophysiology*, 1969, **5**, 597–611.

Grings, W., Lockhart, R., & Dameron, L. Conditioning autonomic responses of mentally subnormal individuals. *Psychological Monographs*, 1962, **76**, (39, Whole No. 558).

Grings, W. W., & Schell, A. M. Magnitude of electrodermal response to a standard stimulus as a function of intensity and proximity of a prior stimulus. *Journal of Comparative and Physiological Psychology*, 1969, **67**, 77–82.

Grings, W. W., & Schell, A. M. Effects of trace versus delay conditioning, ISI variability, and

instructions on UCR diminution. *Journal of Experimental Psychology*, 1971, **90**, 136–140.

KIMBLE, G. A., & OST, J. W. A conditioned inhibitory process in eyelid conditioning. *Journal of Experimental Psychology*, 1961, **61**, 150–156.

KIMMEL, H. D. Inhibition of the unconditioned response in classical conditioning. *Psychological Review*, 1966, **73**, 232–241.

LOCKHART, R. A. Comments regarding multiple response phenomena in long interstimulus interval conditioning. *Psychophysiology*, 1966, **3**, 108–114.

LYKKEN, D. T. Preception in the rat: Autonomic response to shock as a function of length of warning interval. *Science*, 1962, **137**, 665–666.

LYKKEN, D. T. Neuropsychology and psychophysiology in personality research. In E. F. Borgatto & W. W. Lambert (Eds.), *Handbook of personality theory and research*. New York: Rand McNally, 1968.

PROKASY, W. F., & EBEL, H. C. Three components of the classically conditioned GSR in human subjects. *Journal of Experimental Psychology*, 1967, **73**, 247–256.

SLUBICKA, B. Changes in electrodermal responses during classical conditioning and their relationship to personality and awareness. Unpublished doctoral dissertation, University of London, 1971.

SOKOLOV, Y. N. *Perception and the conditioned reflex*. New York: Macmillan, 1963.

STEWART, M. A., WINOKUR, G., STERN, J., GUZE, S., PFEIFFER, E., & HORNUG, F. Adaptation and conditioning of the GSR in psychiatric patients. *Journal of Mental Science*, 1959, **105**, 1102–1111.

(Received July 14, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 318–325

ADDED CUE CONTROL AS A FUNCTION OF REINFORCEMENT PREDICTABILITY[1]

JEROME M. FELDMAN[2]

University of New Mexico

Two experiments were performed in which 88 rats were pretrained to press a lever for food pellets in the presence of an exteroceptive stimulus. Later, a novel stimulus was simultaneously compounded with the pretraining stimulus either all of the time (Exp. I) or some of the time (Exp. II). In Exp. I, the probability of reinforcement in the presence of the compound either continued at the pretraining value or was changed. In Exp. II, the added stimulus was made a more reliable predictor of reinforcement than the pretraining stimulus. Following compound training, the components of the compound were presented individually in a test for stimulus control. Results indicated that only manipulations which altered the probability of reinforcement at the onset of compound training or made the added cue a more reliable predictor of reinforcement than the pretraining cue facilitated added cue control.

Response tendencies which accrue to components of a simultaneous stimulus compound are not always equal in strength. One component may gain more control over behavior than other components, even though all receive equal reinforcement experiences. One of the earliest accounts of this phenomenon was made by Pavlov (1927) during his studies of salivary response conditioning. Pavlov reported that conditioned stimuli which were relatively intense, or which acted on often used sensory receptors, came to control behavior to a greater extent than other stimuli making up the compound. He labeled this effect "overshadowing."

Contemporary interest in overshadowing has led to a modification of the original Pavlovian paradigm. This paradigm, initially demonstrated in research by Lashley (1942), has since been termed "blocking" by Kamin (1968). It has the advantage of not depending on the inherent properties of stimuli, such as intensity or modality, to produce overshadowing effects. Rather, the learning experiences an organism has previously had with stimuli making up the compound determine which of those stimuli will gain control over behavior. Typically, the blocking paradigm first calls for discrimination training along a single stimulus dimension. Later, a novel stimulus dimension is simultaneously compounded with the pretraining dimension as a redundant predictor of reinforcement. After this training, a test is given to assess any control over behavior accrued by the individual components of the compound.

The major result obtained from the blocking paradigm has been that the pretrained stimulus dimension exhibits greater control over behavior, as indicated by subsequent test procedures, than does the added dimension (Chase, 1966; Johnson & Cumming, 1968). In addition, it has been found that added dimension control over behavior is significantly less for pretrained groups than for groups not subjected to pretraining (Kamin, 1968; Mackintosh, 1965a, 1965b). Thus, the order of stimulus learning experiences determines which stimuli come to control behavior. The greater the amount of discrimination pretraining experience, the less likely added dimension control over behavior (Bruner, Matter, & Papanek, 1955; Johnson & Cumming, 1968; Kamin, 1968).

There have been at least three conceptual accounts of blocking results. Selective

[1] This report is based on a dissertation submitted in partial fulfillment of the requirements for the PhD degree at the University of New Mexico. The author is grateful to Frank A. Logan for acting as chairman of the dissertation committee. National Science Foundation grant awarded to Professor Logan supported that dissertation research.

[2] Requests for reprints should be sent to the author, Department of Psychology, Drake University, Des Moines, Iowa 50311.

attention theory (e.g., Mackintosh, 1965c; Sutherland, 1959) has proposed a two-stage process of discrimination learning whereby an organism must first learn to "attend" to a relevant stimulus dimension and then attach appropriate response tendencies to the particular cues making up that dimension. Thus, a stimulus dimension cannot gain control over behavior unless an organism is first attending to that dimension. Since it is also assumed that the probability of attending to one dimension is inversely proportional to the probability of attending to another, blocking results are predictable. Pretraining attention to one stimulus dimension retards attending to, and hence learning about, an added stimulus dimension.

As an alternative to selective attention theory, Kamin (1968) has placed the burden of accounting for stimulus selection results in general, and, blocking results in particular, on an organism's experience with the unconditioned stimulus (UCS) rather than the conditioned stimulus (CS). He suggests that only when the UCS is "surprising" to an organism will an association between the UCS and CS be made. In the blocking paradigm, perception of the added stimulus is assumed to be intact; it is the UCS which is "degraded" at the onset of compound training. The UCS is predicted by the pretraining stimulus, and that prevents added stimulus association because presentation of the UCS is no longer surprising.

A third conceptualization has been advanced by Wagner (1969). The principal assumption of his modified continuity theory is that increments in "signal value strength" to any CS contained in a compound are a function of the degree to which reinforcement is already predictable on the basis of the entire compound. Specifically, the size of this increment is inversely related to the amount of signal value strength already accrued by the compound. In blocking, it is suggested that since the pretraining stimulus has reached asymptotic signal value strength prior to compound training, little or no increments will be accrued by the added stimulus.

Both selective attention theory and modified continuity theory always appear to predict some degree of stimulus selection in the blocking paradigm. However, Kamin (1968) has reported a study in which added stimulus control was observed. Using a conditioned suppression paradigm, it was noted that an increase in the intensity of the UCS (shock) at the onset of compound training, from that which occurred during pretraining, yielded added CS control equivalent to that of the pretraining CS. Kamin (1968) interpreted this finding by suggesting that the change in the UCS was a surprise to the organism and, thus, facilitated added CS association.

It may be possible to extend this concept of UCS surprise into one which includes reinforcement predictability as its major determinant. For example, consider a situation in which the magnitude or intensity of the reinforcing event does not change at the onset of compound training, but the probability of occurrence does. This may be considered a surprising event for the organism and, as such, should facilitate added stimulus control.

While Kamin (1968) may predict that any change in reinforcement probability should facilitate added stimulus control, modified continuity theory and selective attention theory seem to account only for unidirectional effects. For modified continuity theory, an increase in reinforcement probability, presumably raising the limit to which signal value strength can grow, is required. For selective attention theory, a decrease in the probability of reinforcement is required. This would possibly lead to some extinction of the pretraining dimension attention response and, thus, increase the likelihood of attending to the added dimension.

If, as Kamin (1969) suggests, UCS surprise is a requirement for CS association, then the occurrence of an expected or predicted UCS should not facilitate CS association. It may be possible to modify the traditional blocking paradigm so that the added dimension in compound training is a more reliable predictor of reinforcement than the pretraining dimension, without

changing the overall surprise value of the UCS. If, as one might predict, added dimension control is facilitated, possibly as the expense of pretraining dimension control, then Kamin (1969) would not be able to account for such a finding. It should be noted that this experimental manipulation does not fit the traditional definition of "blocking paradigm" because CS redundancy would be absent during the compounding phase. However, it is clearly related to the blocking-experiment class of manipulations and could provide relevant information thereof.

A design employing two experimental variables, change in reinforcement probability and level of added dimension reliability, could be used to assess conditions under which stimulus selection would occur in an appetative conditioning paradigm.

General Method

Subjects

Eighty-eight experimentally naïve male and female hooded rats (90–120 days old) from the University of New Mexico colony served as *Ss* for two experiments. All *Ss* were housed in individual living cages and fed 12-gm. Purina Lab Blox daily. Throughout the experiments, all *Ss* were fed immediately after each training session.

Apparatus

Training and testing were conducted in four noncommercial, sound-attenuated, operant chambers. Each chamber included a houselight, cue light, foodwell, manipulandum, and pellet dispenser. A Hewlett-Packard Model 200B audio oscillator and speaker system were used to produce a 4,000-cps tone at 91 db. during the experiments. Blowers ventilated the chambers and produced a masking noise of 87 db. (re .0002 dynes/cm²) for each chamber.

A 7-w. amber houselight and a white cue light were located 15.24 cm. above the grid floor of each chamber and 5.08 cm. to the right and left of the foodwell, respectively. Intensities of .02 ftl. for the houselight and .50 ftl. for the houselight plus cue light were obtained 7.62 cm. above the manipulanda. Electromechanical relay equipment controlled sequencing and timing of stimuli.

Procedure

Preliminary training.—The *Ss* were randomly assigned to one of four operant chambers and received training in that chamber for the duration of the experiment. On Days 1–4 of training, *Ss* were allowed to make 96 lever presses in the presence of the houselight. Each response delivered one .045-gm. Noyes pellet.

On Days 5–7, one signal (light, tone, or the simultaneous comound of light and tone) was programmed to occur successively through the chambers until a total of 96 trials occurred in each chamber. On Day 5, the signal was presented for 30 sec. in order to increase the probability that a lever press would be made during that interval. The first lever press made in the interval delivered a reinforcement and terminated the signal. If no response was made, the signal terminated automatically at the end of the interval. The average intertrial interval between signal presentations was 35 sec. On Days 6 and 7, maximum signal presentation lengths were 20 and 10 sec., respectively. By Day 8, the first experimental pretraining day, the maximum signal presentation length was 6 sec.

Experiment I

Pretraining.—Thirty-two *Ss* were randomly assigned to four pretraining conditions. Half of the *Ss* were pretrained to respond in the presence of the light and half in the presence of the tone. Half of the *Ss* in each of these groups were assigned to either a 100% or 25% reinforcement probability subgroup. That is, either all of the responses or 25% of the responses (randomly determined) made in the presence of the pretraining cue were reinforced. Pretraining continued for 16 days with 96 possible trials, or cue presentations, given each day.

Compound training.—Sixteen days of compound training, at 96 possible trials per day, followed pretraining. For all *Ss*, the cue not presented during pretraining (light or tone) was simultaneously compounded with the pretraining cue as a redundant predictor of reinforcement. Half of the *Ss* from each of the four pretraining groups maintained their pretraining reinforcement probability during compound training. The remaining *Ss* had their reinforcement probabilities either increased from 25% to 100% or decreased from 100% to 25%. To control for the effects of pretraining experience, two groups of four *Ss* each, without pretraining experience, received compound training for 16 days with either a 100% or 25% reinforcement probability schedule.

Testing.—After compound training, all *Ss* experienced 2 days of component testing. Responses made in the presence of the components were always nonreinforced. To counteract the suspected rapid rate of extinction of responding, test trials were interspersed randomly within 96 regular compound training trials. On Days 1 and 2 of testing, six presentations of the light alone and six presentations of the tone alone were administered.

Results

A performance measure was obtained for each *S* by recording the ratio of responses

made in the presence of a cue to the number of cue presentations given each day. This ratio was then converted to a percentage. Unless otherwise indicated, a .01 two-tailed level of significance was employed to assess differences between groups.

Figure 1 (upper and lower panels) shows the mean percentage made in the presence of a cue, across blocks of days, during pretraining, compound training, and testing. As indicated, during pretraining, groups trained with a 100% reinforcement probability performed better than groups exposed to a 25% reinforcement condition, F (1, 28) = 14.6. However, the performance of all groups improved substantially across blocks of pretraining days, F (7, 196) = 60.2. By the last pretraining block of days, 100% reinforcement groups were responding at a higher asymptote than 25% reinforcement groups, although this difference was not as great as in previous blocks, F (7, 196) = 4.3. Performance differences for groups trained with the light or the tone were minimal and not significant throughout pertraining.

Compound training performance, as indicated by Fig. 1, was similar to terminal pretraining performance in that percentage or responses emitted continued to stay high (about 80–90%). However, groups which were pretrained with the light cue or under a 100% reinforcement condition made a greater percentage of responses to the compound than tone or 25% reinforcement pretraining groups, F (7, 168) = 7.7; F (7, 168) = 2.7, $p < .05$. Compound training performance did not vary as a function of the particular added cue or reinforcement probability that occurred during compound training. Although the performance of pretrained groups did not vary across blocks of compound training days, control group performance did. Percentage of responses increased with increases in compound training, F (7, 42) = 5.2.

In the testing that followed compound training, all groups responded to the pretraining cue presented alone about 80% of the time. Performance did not vary as a function of particular cue or reinforcement

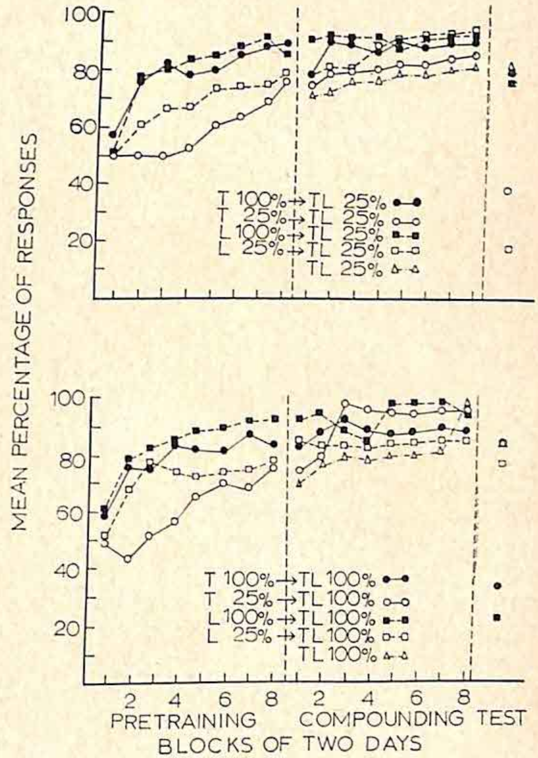

FIG. 1. Mean percentage of responses emitted during pretraining, compound training, and testing in blocks of 2 days. (Test score represent percentage of responses made in presence of added cue. Upper and lower panels indicate performance of groups exposed to 25% or 100% reinforcement probability schedules during compound training.)

conditions experienced during training. Thus, Fig. 1 only shows added cue test data for all groups. In general, for pretrained groups, a greater percentage of responses was made to the pretraining cue than to the added cue, F (1, 24) = 49.6. However, the percentage of responses made to the added cue by pretraining groups was also a function of whether or not reinforcement probability changed at the onset of compound training. Groups which experienced a change made about the same percentage of responses as they did to the pretraining cue. Groups which did not experience a change made significantly fewer responses to the added cue (about 30%) than did groups with a change, F (1, 24) = 40.9. In addition, added cue test performance for pretrained groups with a change in reinforcement did not differ from that of con-

trol groups. Control *S*s responded about 80%–85% of the time to cues making up the compound. Finally, the performance of pretrained groups with a change in reinforcement could not be differentiated on the basis of whether the change was an increase or decrease in reinforcement probability. Under both conditions, *S*s responded about 80% of the time to the pretraining cue and to the added cue.

Experiment II

The results of Exp. I can be predicted from Kamin's (1968) conceptualizations relating CS association and UCS surprise. An added cue gained control over behavior only when the presentation of the UCS was unexpected, that is, when there was a change in reinforcement probability at the onset of compound training.

Experiment II was designed to assess conditions under which added cue control would be exhibited without the concomitant manipulation of UCS surprise. Specifically, the study attempted to ascertain whether or not making the added cue a more reliable predictor of reinforcement than the pretraining cue would also lead to added cue control over behavior. Although not a "true" blocking paradigm, because cue redundancy is not incorporated in the transfer stage of training, Exp. II does have many of the features of typical blocking studies and, as such, may have relevance for the interpretation of their results.

Method

Pretraining.—Two groups of 12 *S*s each were given 10 days of pretraining experience to respond in the presence of a tone or light, respectively. Responses in the presence of the pretraining cue were randomly reinforced 50% of the time. All *S*s were given 96 possible cue presentations each day.

Transfer training.—Following pretraining, 10 days of transfer training, at 96 possible trials per day, were initiated. For all *S*s, the cue that was not present during pretraining was simultaneously compounded with the pretraining cue on those trials in which pretraining cue responses were to be reinforced. On those pretraining cue trials in which responses were never to be reinforced, the added cue was compounded with the pretraining cue either 0%, 25%, or 75% of the time. Thus, for all groups during transfer training, responses in the presence of the pretraining cue were reinforced 50% of the time (the same as during pretraining). However, responses in the presence of the added cue, (always in compound with the pretraining cue) were reinforced either 100%, 80%, or 57% of the time. Responses in the presence of the pretraining cue presented alone were never reinforced.

These manipulations made the added cue for all groups a more reliable predictor of reinforcement than the pretraining cue. At the same time the pretraining cue retained the same 50% reinforcement probability during transfer training as had occurred during pretraining. An additional 24 *S*s, without pretraining experience, were divided into six groups which received transfer training experiences identical to those of the pretrained *S*s. The purpose of these control groups was to enable an assessment of the possible retarding effects of pretraining experience on subsequent transfer training experiences.

Testing.—Test procedures followed the same pattern as that described for Exp. I.

Results

The average percentage of responses emitted in the presence of cues during pretraining, transfer training, and testing is presented in Fig. 2 (three panels). Although pretraining performance by groups exposed to light or tone cues is combined in Fig. 1 (to more clearly show subsequent transfer and test results), and is represented as S_1 50%, *S*s trained with light made a greater percentage of responses than *S*s trained with tone, F (1, 22) = 17.9. However, the performance of all groups improved across blocks of pretraining days, F (4, 88) = 63.7. By the last block of pretraining days, *S*s trained with the tone were responding about 80% of the time; the same as *S*s trained with light, F (4, 88) = 3.4, $p < .05$.

Following pretraining, transfer training was commenced in which an added cue was made a more reliable predictor of reinforcement than the pretraining cue (which maintained a 50% reinforcement probability). Figure 2 shows transfer performance of *S*s in the 100%, 80%, and 57% added cue reliable conditions in top left, top right, and lower left panels, respectively. In addition, transfer performance has been broken down with respect to the percentage of responses emitted in the presence of the pretraining cue alone (S_1 0%

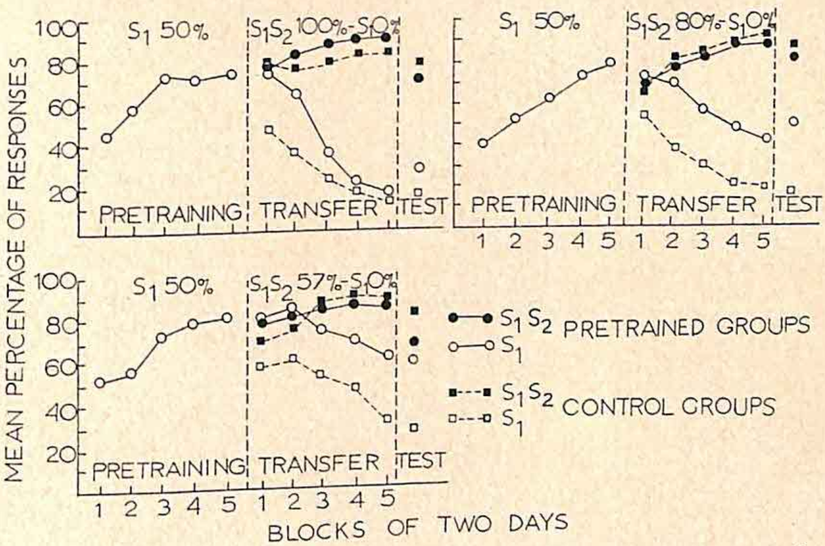

FIG. 2. Mean percentage of responses emitted during pretraining, transfer training, and testing in blocks of 2 days. (In pretraining and transfer training, S_1 represents pretraining cue alone. In transfer training, S_1S_2 represents the compound of added cue plus pretraining cue. Probability of reinforcement in presence of cues is shown as a percentage. Upper left, upper right, and lower left panels depict 100%, 80%, and 57% added cue reliability conditions, respectively. Test scores are percentage of responses made in presence of pretraining cue or added cue alone.)

reinforcement) and the added cue in compound with the pretraining cue (S_1S_2 100%, 80%, or 57% reinforcement).

In general, transfer training performance for all groups improved across training sessions. As training progressed, *Ss* made a greater percentage of responses in the presence of the compound, F (4, 144) = 81.5, and a lesser percentage of responses to the pretraining cue, F (1, 144) = 54.7. Although nonpretrained groups made a lower percentage of responses to the pretraining cue than did pretrained groups, F (1, 36) = 29.0, there was no difference between groups in terms of the percentage of responses made to the compound. Finally, added cue reliability also influenced transfer performance. The more reliable the added cue, the lower the percentage of responses made to the pretraining cue, F (2, 36) = 16.0, and the faster the rate of extinction of responses to the pretraining cue, F (8, 144) = 32.0. The percentage of responses made to the compound did not vary with added cue reliability, although compound performance did improve faster across training sessions for *Ss* having the more reliable added cues, F (8, 144) = 17.5.

Testing, where cues making up the compound were presented individually, essentially mimicked transfer training data. Nonpretrained groups, and groups receiving a high-reliability added cue, performed better in the discrimination than their counterparts, F (1, 36) = 22.4; F (2, 36) = 4.8, $p < .05$. Although nonpretrained groups made about the same percentage of responses (80%) to the added cue as pretrained groups, they made a lower percentage of responses to the pretraining cue than did pretrained groups, F (1, 36) = 4.8, $p < .05$. This difference in pretraining cue responses is accentuated when variations in added cue reliability are considered. The lower the added cue reliability, the greater was the difference in percentage of responses made to the pretraining cue between pretrained and nonpretrained groups, F (2, 36) = 18.1. That is, differences in pretraining cue performance between pretrained and nonpretrained groups were greater for the 57% and 80% added cue

reliability conditions than for the 100% condition.

Discussion

The conceptualizations offered by Kamin (1968; 1969), Mackintosh (1965c), Sutherland (1959), and Wagner (1969) are able to account for those results of Exp. I in which pretraining blocked added cue acquisition and control over behavior. However, only Kamin can fully account for those findings in which pretraining did not retard added cue control, that is, under the conditions of reinforcement probability change in compound training. The basic assumption of Kamin's position is that UCS "surprise" is necessary for CS association. Only when the presentation of a UCS is unexpected will increments in associative strength accrue to a CS which predicts the UCS. In Exp. I it is suggested that groups which maintained their pretraining reinforcement probability during compound training experienced a schedule of reinforcing events that was not surprising to them. Thus, the added cue gained little associative strength and, therefore, little control over behavior. On the other hand, groups which received a reinforcement probability during compound training that was different from the one received during pretraining did experience a surprising schedule of reinforcements. This unexpected schedule presumably enabled associative strength to be gained by the added cue (as well as the pretraining cue) and resulted in added cue control.

It is interesting to note that Kamin's concept of surprise implies that *any* change in a UCS, from that which *S* ordinarily experiences, will facilitate CS association. Although Kamin (1968) has reported that an increase in UCS intensity, shock, did yield added cue control in a conditioned suppression blocking paradigm, he has not reported a study investigating the effects of a decrease in UCS intensity. Experiment I supports the prediction that a decrease in the probability of occurrence of a UCS can also lead to added cue control.

Experiment II was designed to test directly Kamin's assumption that UCS surprise is necessary for CS association. The study employed a *modified* blocking paradigm in which the added cue was a better predictor of reinforcement than the pretraining cue. However, the reliability of the pretraining cue in predicting reinforcement did not change throughout training. Transfer training and test results indicated that the added cue eventually acquired more control over responding than did the pretraining cue. In addition, the greater the reliability of the added cue, relative to the pretraining cue, the faster the rate of discriminative responding during transfer training.

The results of Exp. II are contrary to predictions which can be derived from Kamin's conceptualization. For all pretrained groups during transfer training, there was no overall change in UCS surprise. Probability of reinforcement in the presence of the pretraining cue during transfer training remained identical to that which occurred during pretraining. Yet, added cue control over responding was demonstrated at the expense of pretraining cue control without the concommitant manipulation of UCS surprise.

While transfer training results are predictable from any number of learning theories (e.g., Mackintosh, 1965c; Spence, 1936; Wagner, 1969), it is not readily apparent in the present paradigm how added cue control was acquired at the expense of pretraining cue control. Two factors confound the interpretation of transfer training results. A higher probability of reinforcement occurred in the presence of the compound relative to the pretraining cue alone. Responses in the presence of the pretraining cue alone were consistently nonreinforced and thus extinguished. Despite this acknowledged confounding, results of Exp. II do make one thing clear. UCS surprise, at least the way Kamin has used it, may not always be a necessary requirement for CS association.

Although Exp. II is not a true blocking paradigm (transfer training cues were not redundant in predicting reinforcement), its results do bear on the results of traditional blocking studies. Pretrained animals learned the transfer problem at a significantly slower rate than nonpretrained animals. That is, pretraining retarded, although it did not eliminate, added cue learning. This finding supports predictions made by both the modified continuity theory and the selective attention theory of discrimination learning. The former conceptualization may account for this result by suggesting that pretrained groups first have to reduce the large amount of associative strength accrued by the pretraining cue in pretraining during transfer training. Then associative strength can accrue to an added cue plus pretraining cue. Control groups, however, do not have to reduce pretraining cue associative strength to the degree that pretrained groups must. That is because the

pretraining cue has little associative strength prior to transfer training. Thus, differential control over responding (responding to the compound but not to the pretraining cue alone) occurs sooner for control groups than for pretrained groups. Selective attention theory, on the other hand, can account for these results by making reference to the large degree of extinction training that is necessary to reduce the pretraining cue attention response before the added cue attention response is developed for pretrained groups. For control groups, extinction of a strong pretraining cue attention response is not required because it has not been previously reinforced. Thus, control groups learn the transfer problem at a faster rate than pretrained groups.

In summary, both experiments reported here offer some support for each of the theoretical positions discussed. Compound discrimination learning may be interpreted in terms of an organism's attentional interactions with discriminative stimuli. On the other hand, such learning may be accounted for by stressing an organism's encounter with reinforcing and nonreinforcing events and the degree to which these events are predicted by stimulus compounds as a whole. Whether or not any other conceptualization can more completely and efficiently account for stimulus selection results remains to be determined.

REFERENCES

BRUNNER, J. S., MATTER, J., & PAPANEK, M. L. Breadth of learning as a function of drive level and mechanization. *Psychological Review*, 1955, **62**, 1–10.

CHASE, S. The effects of discrimination training on the development of stimulus control by single dimensions of a stimulus compound. (Doctoral dissertation, City University of New York) Ann Arbor, Mich.: University Microfilms, 1966, No. 66-11-385.

JOHNSON, D. F., & CUMMING, W. W. Some determiners of attention. *Journal of the Experimental Analysis of Behavior*, 1968, **11**, 157–166.

KAMIN, L. J. Attention-like processes in classical conditioning. In M. R. Jones (Ed.), *Miami symposium on the prediction of behavior 1967: Aversive stimulation.* Coral Gables: University of Miami Press, 1968.

KAMIN, L. J. Selective association and conditioning. In W. K. Honig & N. J. Mackintosh (Eds.), *Fundamental issues in associative learning.* Halifax: Dalhousie University Press, 1969.

LASHLEY, K. S. An examination of the continuity theory as applied to discriminative learning. *Journal of General Psychology*, 1942, **26**, 241–265.

MACKINTOSH, N. J. The effect of attention on the slope of generalization gradients. *British Journal of Psychology*, 1965, **56**, 87–93. (a)

MACKINTOSH, N. J. Incidental cue learning in rats. *Quarterly Journal of Experimental Psychology*, 1965, **17**, 292–300. (b)

MACKINTOSH, N. J. Selective attention in animal discrimination learning. *Phycholgical Bulletin*, 1965, **64**, 124–150. (c)

PAVLOV, I. P. *Conditioned reflexes.* London: Oxford University Press, 1927.

SPENCE, K. W. The nature of discrimination learning in animals. *Psychological Review*, 1936, **43**, 427–449.

SUTHERLAND, N. S. Stimulus analysing mechanisms. In, *Proceedings of a symposium on the mechanization of thought processes.* Vol. 2. London: Her Majesty's Stationary Office, 1959.

TRABASSO, T. & BOWER, G. H. *Attention in learning: Theory and research.* New York: Wiley, 1968.

WAGNER, A. R. Stimulus selection and a "Modified Continuity Theory." In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 3. New York: Academic Press, 1969.

(Received July 14, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 326–332

SPATIAL ORIENTATIONAL AND FIGURAL INFORMATION IN FREE RECALL OF VISUAL FIGURES

FRED L. ROYER[1]

Veterans Administration Hospital, Brecksville, Ohio

Three lists of 8 visual figures were constructed by selecting them from subsets of a total set of 63 figures partitioned such that each member of a subset was identical under operations of 90° rotation and mirror reflection. Selection from these subsets, called equivalence sets (ES), produced lists varying according to the number of members of an ES present and the size of the ES from which the members were selected. Each item of the list was exposed for 3 sec.; after a 6-sec. pause at the conclusion of a trial, *S*s freely recalled as many items as possible. Ten trials were given. The number of correctly recalled figures decreased and the number of rotational and reflectional errors increased with the number of members of an ES present in the list and with the size of the ES from which members were selected.

When a task requires *S* to encode a visual figure (hereafter also referred to as a symbol), search an array for that symbol, encode an associated digit, and write out the digit below the symbol, the number of symbol-digit items completed in a fixed time changes as a function of the subset of symbols chosen as stimuli (Royer, 1971). A set of two-dimensional line drawings may be partitioned into subsets that have the property of identity under operations of reflection and 90° rotation; such subsets have been called equivalence sets (ES) by Garner and Clement (1963). As the number of members of an ES selected for a stimulus set increases, the number of symbol-digit items completed in a fixed time decreases. With some special exceptions, the principle holds in the obverse case where *S* encodes a digit, searches an ordered array for the digit, encodes an associated symbol, and writes out the symbol (Royer, 1971). The principle holds also for a card-sorting task where the symbols are matched against criterion symbols (Royer, 1966). In the latter instance there is very little demand on memory; in the former there is. How much might memory have to do with Royer's (1971) findings in the digit-symbol and symbol-digit substitution tasks? The present experiment examines the ability of *S*s to hold stimulus sets of such symbols in memory by testing for free recall.

The importance of this problem lies in previous work, which suggests that there is a two-stage processing of the formal (shape) and spatial orientational information. Garner and Clement (1963) established that the judged goodness of form of figures constructed of five dots in a 3×3 matrix was correlated with the size of the ES (where size is equal to the number of figures in the subset when the total set is partitioned as indicated above). Royer (1966) kept the amount and form of redundancy of subsets of Garner-Clement figures constant in a card-sorting experiment while varying the Gestalten of the subsets. The subsets differed in size of ES from which members were selected and in the number of items from a particular ES. The *S*s used Gestalten, not dot-position information. Performance appeared to be affected by the number of items of an ES present in the stimulus set.

Clement and Varnadoe (1967), following up these observations, had *S*s sort decks containing only two designs in each. The designs in the decks were selected to provide a test of the effects on sorting performance of ES size and ES membership (from same or different ESs). The decks were sorted under two conditions, with orientation either fixed or random; in the latter condition, the items in the deck were rotated randomly so that the figure had to

[1] Requests for reprints should be sent to Fred L. Royer, Psychophysiological Research Laboratory, Veterans Administration Hospital, 10,000 Brecksville Road, Brecksville, Ohio 44141.

be mentally or physically rotated to match the criterion stimulus. There were no differences in sorting times for decks having figures from different ESs whether presented in either fixed or random orientation; however, there were large differences when the figures were from the same ES. Under both conditions sorting times increased as ES size increased.

These observations give strong support to the position that formal and spatial orientational information is processed separately and sequentially. The Clement-Varnadoe (1967) study shows that if formal information alone is the basis of sorting, orientational information has no influence on performance; given a T-like and an F-like figure as criteria, a rotated T-like figure would be as readily discriminated from the F-like figure as a T-like figure in the same orientation as the criterion.

In the type of task Royer (1971) used more recently, rate of processing was faster when each symbol of the stimulus set was from a different ES than when two or more members of an ES were present. However, regardless of the condition, in this task *both* formal and spatial orientational information had to be preserved for *S* to get the item correct.

The differences in the two types of tasks and the similarity of findings raise a new question. In the discrimination tasks such as card sorting, processing formal information alone is possible when criterion stimuli have been selected from different ESs. However, in the symbol-digit substitution task the orientational information still must be processed since it must be preserved in the response. Yet, in the latter task, the same advantage accrues to performance with stimulus sets containing members of different ESs as in the discrimination task. Since the substitution task is more complex in that it requires both discrimination and memory processes, the question is whether the observed results are due primarily to discriminative operations or memory processes that may or may not store information of both types differentially.

The issues may be made more explicit. There are 63 line drawings that can be constructed from the presence or absence of six lines having the spatial orientation shown in Fig. 1a. In constructing lists of eight stimuli by drawing from this set, there are many possible stimulus lists. Regardless of how lists are constructed, each figure in them is unique. However, it is possible to choose the figures according to a previous partitioning of the total set into ESs, as has been done for this experiment; see Fig. 1b. Every figure in each of the three subsets has a particular orientation in space that must be preserved to generate a correct response. In remembering a figure, it is possible that the figure is stored as a complete unit—as a unique figure. Each list has eight figures; if each figure is stored as a complete unit, each figure and each list could be expected to be remembered with equal ease. On the other hand, as in discrimination, it is possible that the information is separately processed and differentially stored. All figures in List A are from different ESs, Lists B and C have fewer items from different ESs. If figural information is more difficult to deal with than orientational, List A ought to be more difficult to remember than B, and B more difficult than C. If, however, orientational information is more difficult to deal with, the opposite order of difficulty would be expected; the uniqueness of the figures depends more on remembering the orientational information. The types of errors that *S*s make should differ according to the treatment of information in memory processes. At the moment, there seems to be no a priori reason to expect that remembering both a T-like and an F-like figure would be more demanding than both a T-like figure and some other figure that is identical to it when rotated 90°.

METHOD

Stimuli.—Three lists of stimuli shown in Fig. 1b were selected from the total set of 63 figures. The first, List A, contains 8 of the 63 figures with each figure being a different item having no other members of its ES present. List B was constructed by choosing two members from each of four ESs having a set size of four members. List C was constructed

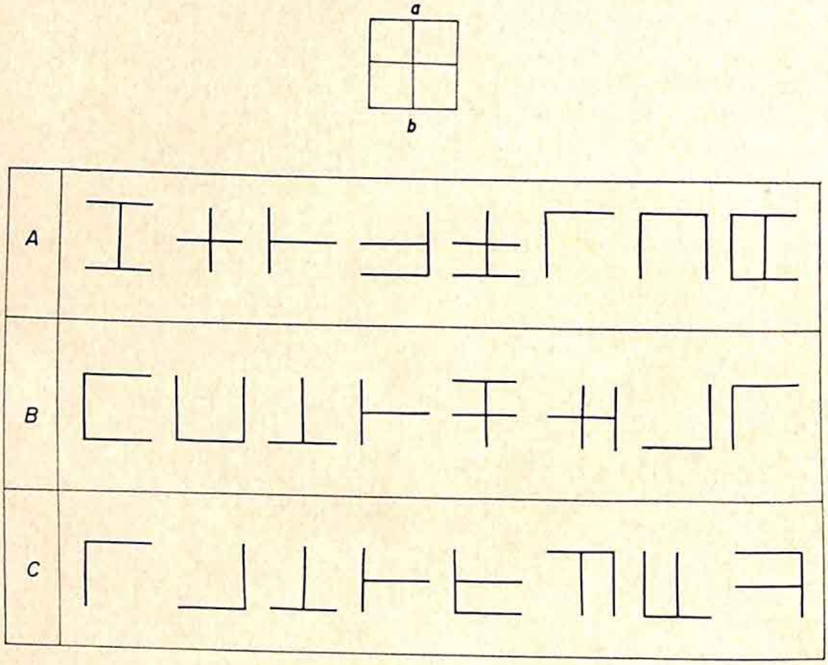

FIG. 1. (*a*) Array of six lines from which the visual figures used in experiment were generated. ((*b*) Lists of figures selected from the total set of 63 line figures generated from the presence or absence of a line in the array shown in (*a*) above.)

by drawing two items from two ESs having a set size of four and by drawing four items from an ES having a size of eight. The selection of a particular item from an ES was determined by the attempt to equate the use of each alternative line position as closely as possible. In contrast to the sets generated for the previous work (Royer, 1966, 1967), complete ESs were not represented since a recognition of the presence of a complete set of equivalent figures would permit *S* to generate items without reference to the spatial orientation information.

Subjects.—Sixty *S*s were obtained from two local academic institutions, Case Western Reserve University and Cuyahoga Community College. Thirty *S*s were obtained from each institution. All *S*s were students in introductory psychology courses. Twenty *S*s served in each of the three experimental groups, defined by the list of figures they were required to learn.

Procedure.—Each of the eight stimulus items of each list was reproduced lithographically on a blank IBM data card. The trimmed corner of the card provided a reference point for the correct spatial orientation of a deck of cards. Each of the eight cards was assigned a position in a deck according to a table of random numbers. An answer card was placed following the eighth card. The answer card consisted of eight lightly printed, dotted-line reproductions of Fig. 1a. Ten packs of the randomly arranged stimulus cards together with an answer card made up the full deck of cards given to each *S*. Decks of each of the three figural lists were both mixed in their order and distributed to *S*s unsystematically so as to approximate randomization.

The experiment was introduced as a study of free recall of geometric designs. The composition of the deck, as 10 sets of eight different designs followed by an answer card, was explained.

The *S*s were instructed that upon the command "Turn," they would turn over a card in the deck and continue turning, only on command, until they had reached a response card. They were told to wait and write nothing on the answer card until instructed to do so. When told, they were to draw on the response card as many designs as they could remember. For illustrative purposes *E* drew a response card on the blackboard, indicating that all designs were constructed by using the six lines on the answer card. It was pointed out that if a figure of a square had been observed as one of the stimulus cards, they would be required to reproduce it from memory by blackening the dotted lines of the perimeter of the figure. The *E* illustrated and told *S*s that (*a*) they would have 3 sec. to observe each stimulus card before being given the command "Turn," (*b*) after turning the eighth card they would wait 6 sec., and (*c*) upon command, they would reproduce all of the items which they could remember. Following these instructions, *S*s were told to remove the rubber band from their deck and position it in front of them. The *E* said "Ready," then gave the command "Turn" and continued at 3-sec. intervals until the eighth stimulus card had been turned over; after a 6-sec. delay, *S* was told to record all of the

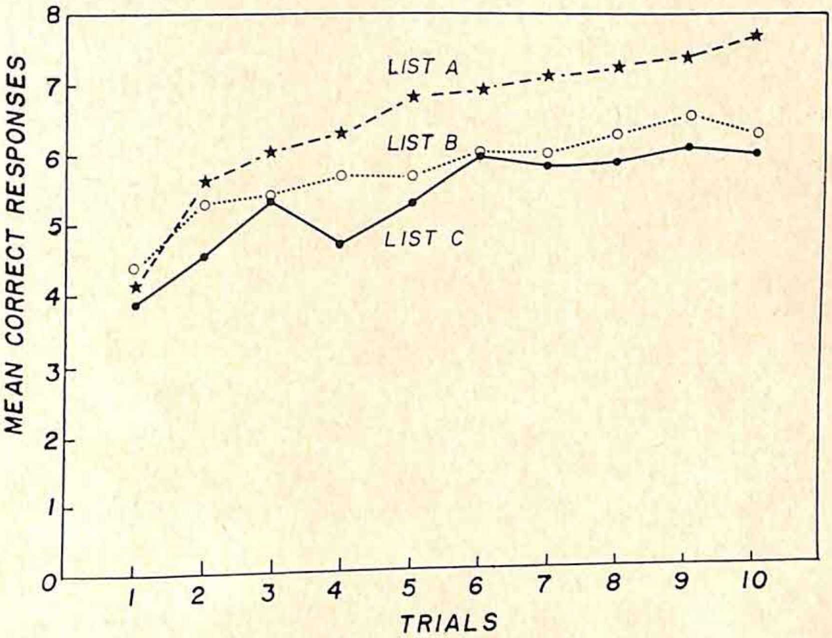

FIG. 2. Mean number of correctly recalled figures for each of the three lists by trials.

items he could remember. This was continued for 10 trials.

Scoring.—For each *S* and each trial the number of correct items and three types of errors were determined. The errors were (*a*) R-R (rotation and reflection), (*b*) Omissions, and (*c*) Others (repetitions and intrusions).

RESULTS

Differences in the free recall learning of the figures.—The mean number of correctly recalled figures for each of the three lists by trials is shown in Fig. 2. The analysis of variance of between-*S*s effects showed that differences in the mean number of figures recalled by the three groups (Lists) are significant, $F\ (2,\ 57) = 7.01,\ p < .01$. Newman-Kuels tests were performed for multiple comparisons with $p < .05$ as an acceptable level of confidence. These tests showed that the means of each of the three treatments (Lists) differed significantly from one another. Within-*S*s effects indicated a significant improvement in performance with trials, $F\ (9,\ 513) = 26.01$, $p < .001$, but even at the end of the tenth trial the differences in learning were still manifest (the Trials $\times$ Lists interaction was not significant, $F\ (18,\ 513) = 1.35$, $p > .05$). In other words, there was no convergence of the learning curves toward equivalent mastery of the lists by the tenth trial. When all items of the list are drawn from different ESs, the list is easier to learn than when there is more than one member from an equivalence set present in the list. Also, the inclusion in the list of one-half of the members of a larger ES (List C) makes the learning more difficult than the inclusion of one-half of the members of two smaller sets (List B). Differences in the ways in which items are selected from a total set of visual figures do influence free recall. The results seem to indicate that the spatial orientational information is fairly well preserved in memory if no other members of an ES are present (List A). However, when several members are present, so that a discrimination of spatial orientation must be maintained in memory, learning is more difficult.

Differences in the types of errors.—The mean number of errors of the three types, (*a*) R-R, (*b*) Omissions, and (*c*) Others (repetitions and intrusions) are presented in Table 1. Intrusion errors were combined with repetition errors since intrusions were

TABLE 1

MEANS AND STANDARD ERRORS OF VARIOUS ERRORS OF RECALL

List	Error type: Rotation and reflection $\bar{X}$	Rotation and reflection SE	Omission $\bar{X}$	Omission SE	Other $\bar{X}$	Other SE
A	7.2	1.12	5.1	1.53	2.1	.48
B	15.6	1.30	6.2	1.46	2.2	.66
C	18.7	1.92	5.3	1.42	3.6	.96

very infrequent and seemed to be only incomplete figures.

The analysis of variance of between-*S*s effects shows that the total number of errors varies significantly with the list, $F\ (2,\ 57) = 6.96$, $p < .01$. Within-*S*s effects show that there are significant differences among the types of errors, $F\ (2, 114) = 66.93$, $p < .001$, and that the types of errors vary with the particular set since the Error Type × Lists interaction is significant, $F\ (4,\ 114) = 6.96$, $p < .01$. Differences in the ways that visual figures are selected from the total set produce differences in the types of errors.

The total errors follow the obverse pattern of the learning scores, of course, since they are complimentary. Because of the significant interactions, Newman-Kuels tests of multiple comparisons were made among means with the acceptable level of significance at $p < .05$. R-R errors increase as the number of members and size of ESs increase; Newman-Kuels tests showed that mean R-R errors for each list differed significantly. There were no significant differences among the means of either Omission or Others types of errors. Within lists, R-R errors were significantly greater than all other types of errors; there were, however, more Omission errors than Others type in both Lists A and B, but no significant difference in these two types of errors in List C. It is seen that the R-R errors account for the major differences in types of errors.

Preservation of formal information.—The data on errors demonstrate clearly that the preservation of spatial orientational information of figural information is affected by the way figures are selected. To examine the question of figural information, we can consider R-R errors as responses preserving figural information under the assumption that *S*s treat members of ESs as nonunique figures. The above analysis of errors suggests an answer, but the answer is clouded by the repetition errors and differential ES repetition.

There is, perhaps, another way of examining the question. We have simply partitioned the maximum possible frequency of response to each ES stimulus item into the numbers of correct reproductions, the numbers of responses which were nonlist members of ESs (R-R), and a residual which represents the nonuse or misuse of presented information. These values were converted into proportions and are given in Table 2. The table also gives the partitioning of proportions for each ES class in each list together with a subpartitioning to show the proportionate use of a particular ES member. The range of the proportions for the residual category for the three lists is from .093 to .114. The small difference of 2.1% seems to leave little doubt that preservation of figural information is not affected by the way that lists are constructed. There is little problem for *S* in remembering the forms regardless of list composition. Remembering which items of an ES are present is a problem however; or, in other words, remembering which alternative spatial orientations are present is a problem.

The proportion of use of different members of ESs varies considerably. Forms which are like common letters or symbols tend to be given as responses with greater frequency. A design labeled 1 would be the common letter or symbol configuration. In List A, this configuration occurs as more than 50% of the nonlist ES items in four of the five possible cases. In List B, the configuration constitutes more than 50% of the errors in all possible cases. In List C, it does not occur in greater frequency than other alternatives in any of the cases.

TABLE 2

PROPORTIONS (P) OF MAXIMAL RESPONSE FREQUENCY FOR LIST, NONLIST, AND RESIDUAL CLASSES OF RESPONSES

List	ES	List items of ES[a]	P list	Nonlist items[b] and proportions of P nonlist	P nonlist[c]	P residual
A	I	1	.865	2(1.00)	.040	.095
	+	1	.945		.000	.055
	T	4	.705	1(.524), 2(.048), 3(.429)	.105	.190
	F	3	.800	1(.161), 2(.000), 4(.161) 1′(.226), 2′(.097), 3′(.355), 4′(.000)	.155	.045
	±	1	.855	2(.220), 3(.778), 4(.000)	.045	.100
	L	2	.750	1(.571), 3(.200), 4(.229)	.175	.075
	U	3	.840	1(.590), 2(.410), 4(.000)	.110	.050
	D	3	.765	1(.526), 2(.368), 4(.105)	.095	.140
			.816		.091	.093
B	U	1, 2	.755	3(.705), 4(.295)	.152	.093
	T	3, 4	.665	1(.746), 2(.254)	.178	.158
	±	3, 4	.742	1(.636), 2(.364)	.192	.065
	L	2, 4	.638	1(.631), 3(.369)	.257	.105
			.700		.195	.105
C	L	2, 4	.672	1(.410), 3(.588)	.255	.073
	T	3, 4	.788	1(.467), 2(.533)	.112	.100
	F	2, 4 1′, 3′	.576	1(.292), 3(.265), 2′(.199), 4′(.243)	.282	.141
			.653		.233	.114

[a] The order of items corresponds to that in Fig. 1, 1b. The numbers in this column indicate the particular item of an ES. The number 1 indicates the design in the reference position given by the design's typographical equivalent in the preceding column. Numbers 2, 3, and 4 indicate the design generated by successive 90° clockwise rotations of Design 1. A prime after a number indicates the design produced by mirror-reflection of Design 1, the number refers to the successive rotation of the 1′.

[b] The numbers indicate the designs of the equivalent sets (ES) as in the above note. In parentheses after the design is the proportion of all rotation-reflection errors for that design.

[c] This value is the proportion of maximum responses which are rotation-reflection errors.

Clearly, there are important exceptions that minimize this as an explanatory principle for memory failure in the experiment. In List C, as noted above, the common configuration is not the most frequent. Also, in List A, for the F-like figure the ES member which is in the normal F configuration occurs 16.1% of the time as a nonlist ES member but the reflection of the stimulus item (mirror-image) occurs 35.5% of the time. And, there are two sets of stimuli which appear in both Lists B and C; the stimuli are from the T-like ES in the three and four configurations and the L-like ES in the two and four configurations. The proportions of the one and two configurations for the T-like figure and of the one and three configurations for the L-like figure in the R-R errors are widely different from one list to another. This emphasizes that responses are not merely reproductions of single stimuli which have been learned or productions of familiar configurations as reversals but are functions of the entire set of items in which they are embedded.

In Lists A and B where items were presented in the common (1 −) configuration, it is of interest that the inverted figure (180° rotation) occurred with greater frequency than the other reversals. Future work may be expected to clarify the specific types of R-R errors associated with the characteristics of the composition of the list.

DISCUSSION

An *S* apparently does not simply remember a figure presented to him as a unique design. As items from the same ES are added to the stimulus set, he begins to show problems of

retaining the figure in memory. The figures are apparently processed as members of sets so that the specification of the unique figure is given separately by the formal information and orientational information. In memory processes, then, the same pattern of results is seen as in discrimination processes. A "figural similarity" is perceived of two-dimensional objects, rotated in space, that satisfy the criterion of identity. The process underlying this perception destroys the uniqueness of the figure.[2] We propose this view of the results. To specify the particular stimulus, a cognitive process is called into operation by which the figure is classified for ES membership and subsequently indexed for orientation. This indexing leads to rotational errors. When only a single item from an ES is indexed, these errors are small, as with List A.

As more items are added from the same ES, as in Lists B and C, the indexing is more difficult or the memory of the index is more difficult to retain; the consequence is that increasing numbers of rotational errors are made although there are fewer alternative ES members available to be wrongly indexed.

It might be emphasized that the poorer performance as ES membership is increased may arise from either poor storage of spatial orientation information or from earlier processing prior to storage such as occurs in the discrimination tasks used in earlier studies. That is, although the task may be a free recall one, discrimination within inferred subsets is required and size and list representation become factors.

Still another possibility is that the information is verbally coded and stored. However, such an explanation would only serve to strengthen the view that antecedent perceptual processes deal with the two types of information separately.

This work emphasizes once more that *S*s do seem to perceive stimuli as members of ESs, even when there are no other members present to require discrimination. Garner (1962) introduced the notion of inferred size of sets to account for the apparent perceived redundancy of a single figure. Garner (1966) has amplified this position in reviewing his own work and that of collaborators. The utility of the concept seems to be demonstrated in the present experiment.

A final note should be made of the clinical aspects of the processing of this type of information. There is abundant evidence in the clinical literature that certain brain damaged cases have considerable difficulty with the type of task used by Royer (1971) and that they tend to produce rotations of visual figures they are asked to reproduce either from memory or during inspection.

Both the substitution task and that of the present experiment require processing of similar information; in one case, the rate of dealing with figures in various rotations is measured, and performance is affected by representation of ES membership; in the other, the number and types of errors are measured, and performance is affected by list representation of ES membership. A population of college students shows performance under certain circumstances that parallels that of brain-damaged individuals. The strong possibility exists that in such patients the damage has interfered with the processes that integrate the separately and sequentially processed information; the interference is manifest under conditions in which it is ordinarily quite rare in normals.

[2] We have elsewhere (Royer, 1971) conjectured that the output of the visual system may be coded only for form with spatial orientation later coded by reference to the body's spatial coordinate information system.

REFERENCES

CLEMENT, D. E., & VARNADOE, K. W. Pattern uncertainty and the discrimination of visual patterns, *Perception & Psychophysics*, 1967, 2, 427–431.

GARNER, W. R. *Uncertainty and structure as psychological concepts.* New York: Wiley, 1962.

GARNER, W. R. To perceive is to know. *American Psychologist*, 1966, **21**, 11–19.

GARNER, W. R., & CLEMENT, D. E. Goodness of pattern and pattern uncertainty, *Journal of Verbal Learning and Verbal Behavior*, 1963, 2, 446–452.

ROYER, F. L. Figural goodness and internal structure in perceptual discrimination, *Perception & Psychophysics*, 1966, 1, 311–314.

ROYER, F. L. Information processing of visual figures in the digit-symbol substitution task, *Journal of Experimental Psychology*, 1971, 87, 335–342.

(Received July 20, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 333–335

SHORT REPORTS

CUE SELECTION AFTER MULTIPLE-CUE PROBABILITY TRAINING[1]

DONALD ROBBINS[2] AND DOUGLAS L. MEDIN

Rockefeller University

Three pairs of letters were presented to *S*s in a two-choice verbal discrimination paradigm. Each pair had an independent probability of being correct (π) and within a pair these values were identical, while between pairs the reward probabilities differed. During training, *S*s were given feedback only with regard to the cue chosen. The training trials were followed by nonrewarded (no information) transfer tests consisting of all combinations of the letters. No preference was observed during training, but the transfer tests revealed a preference for the cue with the higher π value in training. An information model, in which it is assumed that *S*s make inferences about outcomes associated with the unchosen cue of a pair but that these inferred outcomes are less effective than the directly received outcomes, was supported as well as a variant of a "scanning model."

This article deals with an evaluation of a reward information model of choice behavior. The particular task used required *S*s to choose between two simultaneously presented cues, each having a distinct reward probability associated with it. Three such pairs of cues were used and within a pair each cue had the same probability, π, of being called "correct," with π values of .8, .5, and .2 for the three pairs. After training, *S*s were given nonrewarded (no information) test trials on all possible combinations of the six cues.

It is assumed that *S*s assign weights to cues and alter these weights as a result of direct and inferred reward and nonreward. For example, if Cues A and B are presented and if Cue A is chosen and reward follows (*S* is told "correct"), the weight of A changes by:

$$W_{A,n+1} = W_{A,n} + (1 - W_{A,n})\theta_1. \quad [1]$$

If A is chosen and followed by nonreward (*S* is told "incorrect"):

$$W_{A,n+1} = W_{A,n}(1 - \theta_1). \quad [2]$$

If the other cue, Cue B, is chosen and followed by nonreward, the weight of the chosen cue changes according to Equation 2 and the weight of Cue A changes (by the inference that A was correct) according to Equation 1 except θ_1 is replaced by θ_2. Similarly, if Cue B is chosen and followed by reward, the weight of B changes according to Equation 1 and the weight of Cue A changes (by the inference that A was incorrect) according to Equation 2, and again θ_1 is replaced by θ_2.

Note that θ_1 and θ_2 represent the rate of learning on direct and indirect information trials, respectively. These equations describe weight changes following direct reward, direct nonreward, inferred reward, and inferred nonreward. Thus the model is one in which it is assumed that weight changes may differ for inferred and directly received outcomes ($\theta_1 \neq \theta_2$). It is further assumed that these weights are combined to determine choice behavior by a ratio, namely, $P(A/AB) = W_A/(W_A + W_B)$.

Unfortunately, a general expression for the asymptotic weight of a cue as a function of the π values, θ_1 and θ_2, cannot be derived since the application of these operators is partially *S* controlled and a given response probability can be the result of more than just a single set of underlying weights. As a result, a computer simulation was used to obtain average asymptotic weights. For some special cases, one can derive some explicit results. For example, when $\theta_2 = 0$, (i.e., if *S* makes no inference about the unchosen cue), the asymptotic weight of a Cue A approaches π_A. As a result, in the equation determining choice behavior one would substitute π_A and π_B for W_A and W_B, respectively. Results supporting this special case of the model would suggest that *S*s acquire reward information about a cue and it is independent of its context (the π value of the other cue) during training. This model will be referred to as a context-free information case.

When the θ values associated with direct and inferred outcomes are equal, a context-dependent information case results, so that asymptotically $P_A = W_A$. To determine choice behavior, W_A and W_B are replaced by P_A and P_B, the probabilities of choosing Cues A and B, respectively. The context-dependent case would suggest that transfer performance can be predicted directly from the training response probabilities. For many sets of π values (e.g., when the π values for the two cues within a pair are unequal and sum to unity), the context-dependent and context-independent cases cannot be differentiated (see Robbins & Medin, 1970, for a more quantitative treatment). The particular π values used here were selected so as to contrast the three special cases of the model.

[1] This research was supported by Grant GM 16735 from the National Institute of General Medical Sciences. The authors wish to express their appreciation to W. K. Estes for his suggestions and criticisms throughout the course of the research.

[2] Requests for reprints should be sent to Donald Robbins, who is now at Emory University, Department of Psychology, Atlanta, Georgia 30322.

TABLE 1

PREDICTED AND OBSERVED PROPORTION OF A CHOICES (P_A) ON TRANSFER TESTS

Cue A training cond.	Cue B training cond.	Observed P_A	Information model: Dependent case ($\theta_1 = \theta_2$)	Information model: Independent case ($\theta_2 = 0$)	Information model: Partially dependent case ($\theta_1/\theta_2 = 2$)	Cue validity models and scanning models: Model 1	Model 2	3a	3b	3c
.8–.5	*.5*–.5	.59	.50*	.71*	.55	1.00*	.62	.67*	.58	.64
.8–.8	*.2*–.2	.60	.50*	.80*	.60	1.00*	.80*	.85*	.65	.83*
.5–.5	*.2*–.2	.58	.50*	.62	.56	1.00*	.78*	.75*	.58	.73*

Note.—The italicized numerals are the π values for the letters in the transfer tests, while the nonitalicized numerals are the π values for the other member of the training pair. Tests involving the letters from the three pairs were combined and the left-hand member was arbitrarily designated as an A choice.

* $p < .05$, predicted versus observed.

When $\theta_1 \neq \theta_2$, with $\theta_2 \neq 0$, (a partially context-dependent case), the results are intermediate between the completely dependent ($\theta_1 = \theta_2$) and the independent ($\theta_2 = 0$) cases. For example, when $\pi_A = \pi_B = .80$, the context-dependent case yields $W_{A,\infty} = .50$, the context-independent case yields $W_{A,\infty} = .80$, and the partially dependent case yields $.50 < W_{A,\infty} < .80$.

Method.—Twenty-six college age *S*s were obtained in response to advertisements in a local newspaper (*The Village Voice*) and were paid $2.00 each for an approximately 45-min. session.

An 8 × 12 ft. experimental room contained an ASR-33 teletype which was used for presentation of stimuli and outcomes. The teletype was controlled by a PDP-8/I computer housed in an adjacent room. The computer was programmed to control stimulus presentation, response recording, outcome information, and intertrial intervals. The *S*s responded by selecting letters on the keyboard of the teletype. The stimuli were the capital English letters A through F. The six letters were randomly paired to create the three pairings and the assignment of pairs to the π values of .8, .5, or .2 randomized, as was the left–right order of letters. Training consisted of 300 two-choice trials, 100 trials on each of the three pairs. The *S*s were told to choose the letter on each trial which they thought would be most likely to be correct. After each trial, they were told if their choice was correct or incorrect. Thus, if *S* chose A from an A-B pair, the computer would print out "correct" (with probability π_A) or "incorrect" (with probability $1 - \pi_A$). The *S*s were further instructed that they could not be correct all of the time but could improve their performance by paying attention to the results of their choices. Outcomes were printed out immediately after choices with a 1.5-sec. intertrial interval between the outcome printout and the next trial. The pairs were presented in a random order with the restriction that before the $n + 1$st trial on one pair was given the nth trial for each of the other two pairs must have been presented.

Training trials were followed by 30 transfer tests during which outcomes were not printed out. The 30 transfer tests consisted of all of the 15 possible combinations of two letters from the set of six which appeared in both left–right and right–left orders. Prior to the transfer tests, *S*s were told that they were now to choose the letter they thought was most likely to be correct based on what they had just learned and that on these trials they would not be told if they were correct or incorrect.

Results and discussion.—One of the letters from each pair was randomly designated as an A choice. During training, for all pairs, the proportion of A choices oscillated about .50. Examination of the distribution of A choices during the last 60 trials revealed that the response proportions were unimodally distributed around .50 and that individual *S*s did not absorb on one or the other member of a pair.

The results of the transfer tests are shown in Table 1. Tests involving letters with identical training contexts and identical π values were combined. The statistical analyses were *t* tests comparing the observed and expected frequencies of A choices. The major finding is that *S*s showed a statistically significant preference for the higher π value in each of the three transfer tests.

The predictions for the three cases of the information model are also shown in Table 1. Values for the asymptotic weight were obtained by a computer simulation. For each set of π values, 100 statistical *S*s were run for 500 trials each with θ_1 set at .10. The weights for all three π values for the dependent case were all .50, while for the independent case they were .79, .49, and .18 for $\pi = .8$, .5, and .2, respectively.

As shown in Table 1, the dependent case ($\theta_1 = \theta_2$) and the independent case ($\theta_2 = 0$) either overpredicted or underpredicted, respectively, the amount of preference or cue selection obtained. On the other hand, the partially dependent case fared considerably better. Estimates of $\theta_1/\theta_2 = 2$ produced the best fit to the transfer data by a least-squares criterion. The average asymptotic weights obtained in the computer simulation were .59, .49, and .39 for $\pi = .8$, .5, and .2, respectively. Thus these data support a view in which it is assumed that *S*s acquire information about the reward value of a cue and that inferred outcomes are given less weight than directly received outcomes. While the context-dependent case can readily be rejected by these data, conclusions with respect to the context-independent case

may be tempered by the possibility that an insufficient number of training trials were given.

In addition to the information model, some of the cue validity models of Friedman, Rollins, and Padilla (1968) were also evaluated. In these models, the asymptotic validity of a cue is assumed to be equal to (e.g., for Cue A) $\pi_A P_A$ and, since the probability of choosing a cue was approximately .5 at the end of training for all cues, the validity of a cue would be .5 π. This yields values for cue validities or weights of .40, .25, and .10 for π = .8, .5, and .2, respectively. These values were used to generate predictions for the transfer tests. The reader is referred to Friedman et al. (1968) for details of the models. Model 1 assumes that the cue with the highest validity is always chosen, while Model 2 assumes the ratio used in the information models. Models 3a–3c are variants of a "scanning model" (e.g., see Estes, 1966) in which it is assumed that choices are made after *S* generates independent expectancies with regard to a cue being correct or incorrect. Three cases of the model result when the cues are expected to be both correct or both incorrect. Model 3a assumes that new expectations are generated until one in which only one of the cues is expected to be correct is generated; Model 3b assumes that when both cues are expected to be correct or incorrect, the cues are selected with equal probability; Model 3c assumes that if both cues are expected to be correct, then the cues are selected with equal probability and when both cues are expected to be incorrect, new expectations are generated as in 3a.

The predictions are shown in Table 1 and reveals that all but 3b overpredicted the amount of cue selection found. It should be noted that (excluding their "weighted averaging models") Friedman et al. (1968) concluded that 3b and 3c best fit their data and that to more clearly differentiate between the two would require smaller values of cue validities than they had in their study. The present data may meet these requirements.

REFERENCES

Estes, W. K. Transfer of verbal discriminations based on differential reward magnitudes. *Journal of Experimental Psychology*, 1966, **72**, 276–283.

Friedman, M. P., Rollins, H., & Padilla, G. The role of cue validity in stimulus compounding. *Journal of Mathematical Psychology*, 1968, **5**, 300–310.

Robbins, D., & Medin, D. L. Transfer behavior after probabilistic verbal discrimination training. In, *Communications in mathematical psychology*. New York: Rockefeller University, Mathematical Psychology Laboratory, 1970.

(Received January 13, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 336–338

CONTINUOUS TRIAL BETWEEN- AND WITHIN-SUBJECT PARTIAL REINFORCEMENT EFFECT[1]

J. DUTCH[2] AND L. B. BROWN

Victoria University of Wellington, New Zealand

Experiment I compared four groups of rats on resistance to extinction of a bar-pressing response after 800 acquisition trials. In acquisition, Group 1 received continuous reinforcement (CRF), Group 2, fixed ratio 10 (FR 10), and Groups 3 and 4 received both CRF and FR-10 with distinctive stimuli associated with the two acquisition schedules. It was found that partial reinforcement produced a generalizing effect; there was no evidence of stimulus control. Experiment II was identical to Exp. I except that acquisition was continued for 2,900 trials. The generalizing effect of partial reinforcement was again demonstrated, but there was evidence of stimulus control.

As a result of making between-*S* and within-*S* comparisons, discrete trial experiments have shown an effect termed the generalized partial reinforcement effect (GPRE), (e.g., Amsel, Rashotte, & Mackinnon, 1966). The GPRE is found when a within-*S* group is given both partial reinforcement (PR) and continuous reinforcement (CRF) associated with distinctive stimuli in acquisition, and is then extinguished to the CRF stimulus. These *S*s exhibit a partial reinforcement effect (PRE) when their performance is compared with a between-*S* group which has only been given CRF. The Amsel et al. (1966) experiment is particularly important, because when the groups are extinguished the only distinguishing characteristic of the within-*S* and between-*S* groups is their differential acquisition treatment. Their finding of a GPRE is of particular significance for any hypothesis of the PRE, as it suggests that in the discrete-trial situation PR has a primary generalizing influence which seems to preclude the development of stimulus control.

No continuous trial experiment which follows the Amsel et al. (1966) design has yet been published. Pavlick, Carlton, Lehr, and Hendrickson (1967) using continuous trials, obtained a reversed GPRE for their within-*S* group. It should be noted, however, that extinction was conducted on a within-*S* basis, and for the within-*S* group, CRF and PR were alternated in acquisition on a time basis as opposed to the response basis used by Amsel et al. (1966). Pavlick, Carlton, and Manto (1965) did obtain a within-*S* PRE, but in this experiment there were no between-*S* groups and extinction was conducted on a within-*S* basis.

The following experiments were designed to determine whether, in the continuous trial bar-pressing situation, the use of a response-based schedule switch will result in a GPRE, thus indicating continuity between runway and continuous trial effects. Four groups were used in each experiment: two within-*S* PR-CRF groups, a between-*S* CRF control, and a between-*S* PR control. If stimulus control is a critical feature of the PRE, it would be predicted that within-*S* CRF *S*s should extinguish like between-*S* CRF *S*s. If on the other hand a GPRE were found, it would demonstrate continuity between the PRE in the discrete trial and continuous trial situations and influence of PR independent of stimulus control.

Method.—The *S*s were 80, experimentally naïve male Sprague-Dawley rats from the Victoria University of Wellington colony. Ages ranged from 100 to 120 days at the commencement of the experiment.

The apparatus consisted of two Lehigh Valley Electronics, Inc. small animal test chambers, each enclosed in a Lehigh Valley Electronics sound-insulated cubicle equipped with a blower. One of the two bars in each box was removed and the remaining bar was set in the high position, 16.5 cm. above the grid floor. The bar required an operating pressure of 24 gm. Reward consisted of .01 cc. of water dispensed by a liquid dipper. Particular stimuli were associated with reinforcement. These consisted of a white house light situated in the sound-insulating cubicle at the center of the rear wall and a red cue light situated directly above the operating bar.

Three weeks before the commencement of the experiment, *S*s were placed on a 23.5-hr. water deprivation schedule with ad-lib food. Magazine training was conducted over 3 days and consisted of 30 min. per day in the apparatus with the bar absent and the dipper programmed to operate once every 30 sec. on the average. On the fourth day, the bar was inserted and *S* was required to make 100 responses on CRF. Acquisition commenced on the following day.

Particular stimuli were associated with reinforcement. For Groups 1 and 2, the red and white lights alternated every 50 trials. In the within-*S* groups, 3 and 4, half of the *S*s had white light in conjunction with the CRF and red light in conjunction with PR, while for the other half this relation was reversed. The daily acquisition sessions were broken into 50 trial blocks, with no time space between the blocks. At the end of each block, the schedule and its accompanying cue light changed. The presentation of blocks was randomized, with the restriction that the same condition could not prevail on two consecutive blocks of trials.

[1] This research was aided by Grant 67/68 from the New Zealand University Grants Committee to the first author.

[2] Requests for reprints should be sent to J. Dutch, Department of Psychology, Victoria University of Wellington, P. O. Box 196, Wellington, New Zealand.

On Day 1 of acquisition, all *Ss* received a total of 200 trials; for the remainder of acquisition, all *Ss* received a total of 300 trials/day. The PR component on Day 1 was FR-2; on Day 2 it was FR-5; on Day 3 and all subsequent days it was FR-10. The *Ss* in Exp. I received a total of 800 acquisition trials. In Exp. II, acquisition was continued until all *Ss* received a total of 2,900 trials.

In a pilot study, it was found that the changes in ratio over the 3 days resulted in a large number of within-*S Ss* failing to complete acquisition. However, it was thought necessary to persist with these ratios to maximize the PR effects. In order to minimize the possible effects of bias through *Ss* failing to complete acquisition, the following procedure was used in both experiments. At the commencement of acquisition, two pools were made up consisting of 10 *Ss* each. These two pools were later used to make up Groups 3 and 4. Pool 1 consisted of 10 *Ss* assigned to White CRF-Red PR. Pool 2 consisted of 10 *Ss* assigned to Red CRF and White PR. At the end of acquisition, 5 *Ss* from each of these two pools were randomly selected to make up Group 3, the remainder constituting Group 4. During the course of acquisition, 4 *Ss* were discarded from Exp. I, and 5 *Ss* were discarded from Exp. II. These *Ss* were discarded for failing to complete one block of trials within 2 hr. Of the 4 *Ss* which were discarded in Exp. I, 3 were from Pool 1 and the other was from Pool 2. Of the 5 *Ss* discarded in Exp. II, 3 were from Pool 1 and the other 2 were from Pool 2.

Extinction followed on the day after the completion of acquisition. The criterion for extinction was 10 consecutive min. of nonresponding.

Half of the *Ss* in Groups 1 and 2 were extinguished under white light, the other half, under red light. All the *Ss* in Group 3 were extinguished under stimuli associated with CRF, and all the *Ss* in Group 4 were extinguished under stimuli associated with PR, so that in each group half of the *Ss* were extinguished under white light and the other half, under red light. The stimuli were not changed during the extinction session, so that for all *Ss* extinction was conducted on a between-*S* basis.

Results and discussion.—Table 1 shows the mean number of responses in extinction for each group in Exp. I and II.

TABLE 1

Mean Number of Responses in Extinction

Group	Exp. I	Exp. II
1 (CRF between)	59.1	68.1
2 (PR between)	147.1	348.8
3 (CRF within)	107.1	286.9
4 (PR within)	102.8	392.4

For Exp. I, a comparison was made between all groups on resistance to extinction using the Kruskal-Wallis one-way analysis of variance by ranks, with $H\ (3) = 17.644,\ p < .001$.

Group 1 (CRF between) was less resistant to extinction than Group 2 (PR between), $U = 3$, $p < .001$; Group 3 (CRF within), $U = 9.5, p < .001$; and Group 4 (PR within), $U = 17, p < .01$. Group 2 was more resistant to extinction than Group 4, $U = 27,\ p < .05$. There was no significant difference between the two within-*S* groups, Groups 3 and 4, $U = 46,\ p > .05$.

For Exp. II, a comparison was made between all groups on resistance to extinction using the Kruskal-Wallis one-way analysis of variance by ranks, $H\ (3) = 24,\ p < .001$.

Group 1 (CRF between) was again less resistant to extinction than Group 2 (PR between), $U = 0$, $p < .001$; Group 3 (CRF within), $U = 0, p < .001$; and Group 4 (PR within), $U = 0, p < .001$. There was not a significant difference between Groups 2 and 3, or between Groups 2 and 4. Group 3 was less resistant to extinction than Group 4 ($U = 21$, $p < .05$).

In both experiments a PRE was obtained, Group 2 being more resistant to extinction than Group 1. A GPRE was also obtained in both experiments, Groups 3 and 4 being more resistant to extinction than Group 1. These findings are important in that they establish concordance between discrete trial and continuous trial partial reinforcement phenomena. The results of Exp. II do, however, differ from Exp. I in that a within-*S* PRE has been obtained in Exp. II, Group 4 being more resistant to extinction than Group 3.

In the continuous trial situation, resistance to extinction is a function of the absolute number of nonreinforced trials (Dutch and Quartermain, 1967). This suggests that "persistence," defined as a specific energy or work requirement, is necessary for PR to be effective in obtaining a PRE. Further support for this element can be indirectly derived from the experiment of Pavlik et al. (1967). They obtained a reversed PRE for within-*S* groups when their acquisition and extinction schedule changes were time based. Taken in conjunction with the findings from Exp. I and II, it is assumed that persistence generalizes to the bar-pressing response as a whole and the strength of this effect on behavior is such that it is difficult to bring persistence of the response under stimulus control, hence the PRE and the GPRE.

The hypothesis best suited to accommodate these findings is that of Amsel et al. (1966). There must, however, be an additional factor operating in the continuous trial situation which permits the development of a within-*S* effect. It seems likely that this additional factor was the possibility of developing different modes of responding in the continuous trial situation. For example, in the continuous trial situation, an *S* on an FR-10 schedule tends to make a chain of responses approximating to the length of the schedule, in this case a run of approximately 10 responses (e.g., Bindra, 1963), whereas on a CRF schedule the responses are clearly single, with each response being separated by consummatory behavior. Exposing *Ss* to the two schedules in the within-*S* design therefore developed a different mode of response for each schedule. In this way, dis-

tinctive modes of response became associated with the distinctive stimuli accompanying the distinctive schedules of reinforcement (e.g., Ross, 1964).

REFERENCES

AMSEL, A., RASHOTTE, M. E., & MACKINNON, J. R. Partial reinforcement effects within subject and between subjects. *Psychological Monographs*, 1966, **80**, (20, Whole No. 628).

BINDRA, D. Temporal analysis of relevant and irrelevant behavior components in partial reinforcement and extinction. *Psychological Reports*, 1963, **13**, 551–563.

DUTCH, J., & QUARTERMAIN, D. Unrewarded trials and resistance to extinction of a bar pressing response. *Psychonomic Science*, 1967, **9**, 505–506.

PAVLIK, W. B., CARLTON, P. L., LEHR, R., & HENDRICKSON, C. A reversed PRE. *Journal of Experimental Psychology*, 1967, **75**, 274–276.

PAVLIK, W. B., CARLTON, P. L., & MANTO, P. G. A further study of the partial reinforcement effect within-subjects. *Psychonomic Science*, 1965, **8**, 533–534.

ROSS, R. R. Positive and negative partial reinforcement extinction effects carried through continuous reinforcement, changed motivation, and changed response. *Journal of Experimental Psychology*, 1964, **68**, 492–502.

(Received March 10, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 338–340

ENCODING PROCESSES IN THE STORAGE AND RETRIEVAL OF SENTENCES

RICHARD C. ANDERSON[1]

University of Illinois

In two experiments people attempted to recall the subject noun of each of a series of once exposed sentences given the verbatim predicate or its paraphrase. Groups that received imagery instructions recalled two and one-half times as many words as a group that received oral repetition instructions, yet the conditional probability of correct recall on a paraphrase item given correct recall on the verbatim item, based on the same sentence, was nearly as high in the repetition group (.83) as it was in the imagery groups (.92). The implication is that the storage of sentences usually entails semantic encoding. Nonetheless, performance on verbatim items was consistently somewhat better than performance on paraphrase items, a fact not attributable to recall from a short-term, phonological store and probably not attributable to inexactness of paraphrasing.

Bobrow (1970) has argued that when people learn from sentences, they store meanings rather than strings of speech sounds. He presented pairs of nouns twice, each time embedded in a sentence (e.g., *The animal's bark scared the big league pitcher*). Changes in the second sentence did not affect recall of the last noun in a sentence, given the first noun as the retrieval cue, as long as the referents of the nouns remained the same (e.g., *The dog's bark frightened the baseball pitcher*). However, when the sense of the nouns changed (e.g., *The medicinal bark filled the porcelain pitcher*), recall was substantially lower. If *S*s were storing phonological codes, context-induced changes in the meanings of the words would not have affected performance. Hence, the results imply that meanings were stored rather than sounds.

The present study used a different strategy to explore further whether the long-term storage of sentences entails sounds or meanings. After a single presentation of a list of sentences, *S*s attempted to recall the subject nouns given as retrieval cues verbatim and paraphrased predicates. The logic was that if a person had stored *only* a phonological code, he would be able to answer a verbatim item, but unable to answer the paraphrase item based on the same sentence. If a person could answer the verbatim item and the paraphrase item for a sentence, then he must have stored a semantic code or have stored both a semantic and a phonological code.

One objection to Bobrow's (1970) study is that all his *S*s composed continuation sentences during learning, a procedure that presumably makes meaningful processing more probable. It may be that people can learn and remember a phonological code if conditions predispose them to do so. To investigate this possibility, half of the *S*s in the present experiment were instructed to repeat each sentence aloud again and again, a procedure that should make phonological encoding more likely. The remaining *S*s were instructed to form a vivid mental image of the event described in each sentence. Imagery instructions are known to facilitate sentence learning (e.g., Anderson & Hidde, 1971), and forming an image surely entails semantic encoding.

Method.—1. Experiment I: The *S*s were 48 undergraduates, participating to fulfill an educational psychology course requirement and randomly assigned to conditions upon appearance for the experiment.

Twenty pairs of sentences were constructed. Within each pair, the subject nouns were identical

[1] Requests for reprints should be sent to Richard C. Anderson, Department of Psychology, Training Research Laboratory, University of Illinois, 8 Lincoln Hall, Urbana, Illinois 61801.

while the predicates were, in the judgment of *E*, fairly close paraphrases of one another. Most of the sentences consisted of a subject, a verb, and a direct object in that order. Except for the subject noun and articles, all the words in a sentence and its paraphrase were different. The subject nouns designated kinds of people (e.g., occupational titles). They were selected from among such nouns with a frequency of between 200 and 700 occurrences per $4\frac{1}{2}$ million in the Thorndike-Lorge List T. The nouns were assigned at random to sentence pairs, with the restriction that there was not an obvious association between the noun and the predicates. Some examples of pairs of sentences are: *The governor was mad at the instructor* and *The governor was angry with the teacher; The uncle shouted an obscene remark* and *The uncle yelled some dirty words; The traveler appreciated the gift* and *The traveler was grateful for the present.*

Half of the *S*s were instructed to bring to mind a vivid mental image of the event described in each sentence. The remainder were told to read each sentence aloud as many times as they could during the presentation interval. During learning, half of the *S*s received one sentence from each pair; the other half received the remaining sentence. There were two forms of the test, and every *S* got both forms in counterbalanced orders. Half of the items in each form of the test repeated the predicate of a presented sentence and half paraphrased the predicate of a presented sentence. Over the course of the two tests, every *S* received both a verbatim and a paraphrase item for each sentence.

The *E* presented sentences typed on 3×5 in. file cards one at a time at a 6-sec. rate. The presentation was paced by beeps from a tape recorder. Following one trial on the 20 sentences, the two tests were administered. The test items, which were also typed on 3×5 in. file cards, consisted of the sentences with $\frac{7}{8}$-in. blanks in place of the subject nouns. The tests were *S* paced. Oral responses were required.

Following Nelson's (1970) procedure for insuring that only items in long-term storage are available at recall, sentences and test stimuli were presented in blocks of 10 items. The block of items presented first during learning also appeared first during each test. The order of blocks was counterbalanced. The forms of the test were arranged so that half of the items within each block were verbatim and half were paraphrased. Within blocks the order of sentences or test items was randomized by shuffling the file cards. The blocking procedure usually introduced an interval of at least 1 min. between presentation and testing. However, since an *S*-paced testing procedure was employed, the retention interval was less than 1 min. on some items for some *S*s.

2. Experiment II: The *S*s were 37 undergraduates who participated to fulfill a requirement in an introductory educational psychology course. Forty-eight *S*s were to have been involved; however, the experiment was terminated early because the pool of available *S*s was exhausted. To compensate for the fact that there were disproportionate numbers of cases per cell, an unweighted means analysis of variance was employed.

All *S*s received imagery instructions. Following one presentation of the sentences, the same ones used in Exp. I, half of the *S*s completed arithmetic problems for 2 min. before receiving the tests. The remainder took the tests immediately. Unlike Exp. I, the test items were presented at a 6-sec. rate. The arithmetic problems—pairs of two-digit numbers to be added or subtracted—were also presented at a 6-sec. rate. In all other respects the two studies were identical.

Results.—1. Experiment I: The imagery group recalled 65.2% of the subject nouns, whereas the repetition group recalled only 24.9%, F (1, 46) $= 52.39$, $p < .01$. The *S*s recalled 46.6% given a verbatim cue and 43.5% given a paraphrase cue, a small but significant difference, F (1, 46) $= 8.42$, $p < .01$. The only other significant effect was the Type of Test $\times$ Test Order interaction, F (1, 46) $= 4.67$, $p < .05$; recall was greater for verbatim than paraphrase items on the first test but slightly greater for paraphrase items on the second test.

For the most part, *S*s either got both the verbatim and the paraphrase items for a given sentence right or they got both items wrong. The proportion of both right or neither right was .925 in the imagery group and .940 in the repetition group. Pooling the data from the imagery and repetition groups, which were negligibly different, verbatim items were correct alone .049 of the time while paraphrase items were correct alone in .019 of the cases. The conditional probability of a correct response to a paraphrase item given a correct response to the verbatim item based on the same sentence was .922 in the imagery group and .827 in the repetition group, t (46) $= 1.60$, $p > .05$.

2. Experiment II: Recall of subject nouns averaged 68.1% given a verbatim cue and 63.5% given a paraphrase, F (1, 29) $= 25.00$, $p < .01$. Recall was slightly, but not significantly, lower after a 2-min. delay, $F < 1$. Presence or absence of the delay was not a factor in any interaction. Once again, paraphrase performance improved slightly from the first to the second test whereas verbatim performance got somewhat worse, but in this experiment the Type of Test $\times$ Test Order interaction was not significant, $F < 1$.

The proportion of cases in which the verbatim and paraphrase items for a given sentence were either both right or both wrong was .931. As in Exp. I, when exactly one item was correct it was more often the verbatim item than the paraphrase. Verbatim and paraphrase items were correct alone in .058 and .011 of the cases, respectively. The conditional probability of recall on a paraphrase item given recall on the verbatim item based on the same sentence was .915.

Discussion.—The present studies demonstrate that when people are able to answer a verbatim question from memory, they are almost sure to be able to answer a paraphrase question based on the same sentence. This is strong evidence that learning from

a sentence usually entails semantic encoding, because a paraphrase and its base are related in terms of meaning but unrelated with respect to the shape or the sound of the words.

Pooling across all groups in both experiments, there were 163 cases in which *S* made an overt error on both the verbatim and paraphrase items for a sentence. In 66.9% of these cases, the error word was the same on both items. This fact indicates that even erroneous information is highly accessible to a semantic probe and is further evidence for the importance of semantic encoding in long-term storage.

In Exp. I, some *S*s received imagery instructions, which should induce semantic encoding. Others received oral repetition instructions, which could be expected to induce only phonological encoding, if indeed this is possible. Recall averaged better than two and one-half times as high in the group that received imagery, yet the conditional probability of being right on a paraphrase item given a correct response to the verbatim item based on the same sentence was nearly as high in the repetition group as it was in the imagery group. These data appear to make untenable the position that phonological and semantic encoding give rise to forms of long-term storage which can be used alternatively, depending upon *S*'s strategy or circumstances.

Notwithstanding the strong correlation between verbatim and paraphrase performance, recall was consistently somewhat better on verbatim than paraphrase items. There are several possible explanations of this fact. It seemed possible that the superior performance in Exp. I on verbatim items included in the first test was attributable to recall from a short-term phonological store. However, in Exp. II superior recall of verbatim items appeared even after a filled interval of 2 min.

Another possibility is that the difference between verbatim and paraphrase recall appeared because some of the paraphrases were inexact. A rough paraphrase might fail to induce a successful search for a stored semantic representation. To check this possibility, 14 *S*s rated the pairs of sentences on a 5-point scale ranging from "nearly identical in meaning" to "rather different in meaning." The reliability of the average rating, based on the intercorrelations among ratings, was .85. However, in Exp. I the correlation between the average ratings and the item by item differences between verbatim recall and paraphrase recall was − .13, meaning that there was a slight tendency for exact paraphrase to be associated with poor recall. In Exp. II the correlation between the average ratings of paraphrase exactness and item-by-item differences between verbatim and paraphrase recall was .04. The correlation between the item by item differences from the two studies was actually negative, − .07. Evidently the inferior performance on paraphrase items cannot be accounted for in terms of roughness of paraphrasing.

Based on the data at hand, the conclusion seems to be that while the long-term storage of sentences *usually* entails semantic encoding, surface orthographic or phonological information is encoded at least some of the time and can be sufficient for recall from the long-term store.

REFERENCES

Anderson, R. C., & Hidde, J. L. Imagery and sentence learning. *Journal of Educational Psychology*, 1971, **62**, 526–530.

Bobrow, S. A. Memory for words in sentences. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 363–372.

Nelson, T. O. An investigation of the forgetting process in long-term memory: Benefits of relearning over original learning. Unpublished doctoral dissertation, University of Illinois, 1970.

(Received March 22, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 341–343

SEQUENTIAL AND NONSEQUENTIAL MEMORY FOR VERBAL AND NONVERBAL AUDITORY STIMULI[1]

RONALD P. PHILIPCHALK[2] AND EDWARD J. ROWE

University of Western Ontario

Lists of 20 easily recognizable sounds or their corresponding verbal labels were presented for five trials under either free or serial recall conditions. Correct recall of the two types of items did not differ in free recall, but the words were recalled significantly better than the sounds in serial recall, thus showing that verbal material is more readily recalled than nonverbal when the correct serial order of items must be retained. The findings were discussed in terms of A. Paivio's two-process theory of memory.

This study was designed to investigate memory for meaningful nonverbal auditory stimuli, i.e., familiar sounds. Despite a recent interest in the psychological processes underlying memory for nonverbal visual material (i.e., pictures and drawings of familiar objects), little information is available on the factors which affect learning and retention of meaningful sounds, the auditory counterpart of visual pictures. A recent study by Miller and Tanis (1971) has demonstrated that familiar sounds are as easy to recognize as spoken words after a single presentation, although the level of performance is below that found for pictures by other investigators. The present study represents a further attempt to determine the nature of the factors affecting memory for auditory stimuli.

Two comparisons were made in the experiment, i.e., memory for both sounds and their corresponding verbal labels was studied as a function of sequential (serial recall) versus nonsequential (free recall) task requirements. The first comparison was designed simply to determine whether, in fact, any differences in recall exist between words and sounds such as have been found for words and pictures in the visual modality (e.g., Sampson, 1970). The free recall (FR) – serial recall (SR) comparison was prompted by the results of recent studies showing that under certain conditions, memory for verbal material is superior to both visual (Paivio & Csapo, 1969) and auditory (Warren, Obusek, Farmer, & Warren, 1969) nonverbal material when the stimulus items are to be recalled in a specific order. Thus, the sequential aspect of the recall task might be one characteristic which could serve to differentiate memory for material that is verbal or nonverbal in nature.

Method.—The *S*s were 80 introductory psychology students who participated as part of a course requirement. They were assigned to one of four experimental conditions in groups of from 3 to 7, with the number of *S*s in each condition equated at 20.

A list of 20 sounds was selected from a larger pool of 67 items, most of which had been obtained from commercial sound-effects recordings. As part of a previous study,[3] these 67 sounds were played to a group of 33 students of introductory psychology, who were required to supply a verbal label for each. The items used in the present experiment were those labeled consistently by at least 88% ($\bar{X}$ = 95.4) of the normative group, and ranged in length of presentation time from 3 to 6.5 sec. ($\bar{X}$ = 5.4). The labels of the sounds were COUGH, DOOR, TRAIN, BELL, TYPEWRITER, DRUM, TELEPHONE, CAT, ROOSTER, CLOCK, HORSE, LAUGHTER, DOG, HORN, SAW, HAMMER, APPLAUSE, SIREN, CYMBALS, and WIND.

The sounds were recorded on tape in the above order with an interitem interval of 1 sec. This list was then re-recorded five times on another tape to produce the stimuli for five SR trials. Five different random orders of the items were used in FR. The verbal list consisted of the labels of the 20 sounds, recorded on tape by a male voice at a 6-sec. rate. The presentation rate for the two types of material was thus approximately equal, as the corresponding rate for the sounds averaged 6.4 sec.

Conventional instructions for either FR or SR were read by *E* at the beginning of each experimental session. The *S*s in the verbal groups were instructed to write down as many words as they could remember after each trial, either in any order (FR) or in the order in which they were presented (SR). The *S*s in the sound groups were similarly instructed, except that they were told to "try to use a single word which could be used to label each sound" in recording their responses. One minute was allowed for recall in all groups. The answer sheets for each trial were collected at the end of the recall period.

Results and discussion.—The recall sheets were scored by both a strict and lenient scoring procedure, the difference between the two being that apparent auditory confusions in the case of the word lists (e.g., CLOTH for COUGH) and questionable labeling of the sounds (e.g., PARTY for LAUGHTER) were scored as correct only by the lenient criterion. Homophones (e.g., SYMBOLS for CYMBALS) and synonymous labels (e.g., CLAPPING for APPLAUSE) were counted as correct in both procedures. The analysis of correct recall for the two scoring criteria yielded highly

[1] This research was supported by a grant from the National Research Council of Canada (0087) to A. Paivio and a Postgraduate Scholarship to the second author from the National Research Council of Canada. Thanks are due to Al Paivio for helpful criticisms at all stages of the research.

[2] Requests for reprints should be sent to R. P. Philipchalk, Department of Psychology, St. Thomas University, Fredericton, New Brunswick, Canada.

[3] This study was conducted in collaboration with Carole H. Ernest.

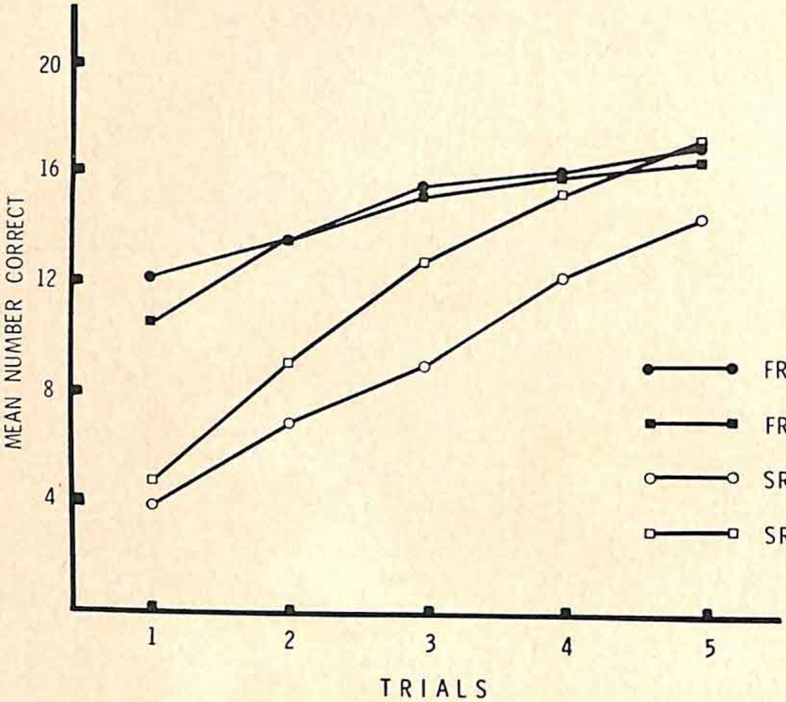

FIG. 1. Mean number correct for sounds (S) and words (W) in free recall (FR) and serial recall (SR) as a function of trials.

similar results, and only those for the strict procedure will be presented here.

The number of items correctly recalled in each condition are shown in Fig. 1. The data were analyzed by a $2 \times 2 \times 5$ analysis of variance with stimulus class (words and sounds), recall task (FR and SR), and trials as factors. The main effect of task was significant ($p < .001$), with higher recall occuring for the FR groups, as was the effect of trials ($p < .001$) and the Task $\times$ Trials interaction ($p < .001$). The interaction was due to a more rapid increase in recall scores across trials for the SR group. The main effect of stimulus class was not significant, but this factor interacted with task, $F(1, 76) = 8.95$, $p < .01$, such that words were recalled better than sounds in SR but the two did not differ in FR. With the data collapsed across trials, post hoc comparisons using two-tailed t tests showed that the difference between sounds and words was significant in SR, $t(38) = 3.40$, $p < .01$, but not in FR, $t < 1$. The results show that under the present experimental conditions, memory for sounds is affected to a greater degree by the sequential nature of the SR task than is memory for words. As a further test of this conclusion, the data for SR were rescored for correct recall only, i.e., ignoring incorrect serial position. The mean recall per trial was 14.06 for words and 13.11 for sounds, a difference which falls short of significance, $F(1, 38) = 2.82$, $p > .10$. Thus it seems reasonable to conclude that the difference for SR illustrated in Fig. 1 is the result of the sequential recall requirements, rather than some more general factor such as decreased availability of the sounds in this condition.

The present findings form an interesting parallel to the results of Warren et al. (1969). These experimenters found that *S*s generally exhibited an inability to remember the correct order of four sounds (a high tone, a low tone, a hiss, and a buzz) presented in rapid succession at a duration of 200 msec. each. On the other hand, correct serial recall of four spoken digits presented under the same conditions produced little or no difficulty, thus indicating that order information is more readily available for verbal material. Comparison of these results with those of the present study is somewhat tenuous because of the different experimental conditions (e.g., the time parameters involved), but the two sets of findings are consistent in suggesting that auditorially presented verbal material is recalled better than nonverbal when the correct ordering of items must be maintained. The results for the FR condition alone are also comparable to those of Miller and Tanis (1971), who found no difference between familiar sounds and spoken words in recognition memory.

The results may be interpreted theoretically in terms of a two-process theory of memory proposed by Paivio (1971), which postulates the existence of both verbal and nonverbal (visual imagery) memory codes. These coding systems may be differentiated functionally by their relative specialization for sequential and parallel processing of verbal and nonverbal information, respectively. Paivio and

Csapo (1969) have presented evidence that performance in nonsequential memory tasks (free recall, recognition memory) varies with the availability of both memory codes, while the verbal code alone is crucial in sequential memory (immediate memory span, serial learning). The availability of imaginal and verbal codes was varied by these *E*s through the manipulation of presentation rates and stimulus concreteness, with the stimuli being presented visually. Their results strongly supported the postulated distinction between verbal and nonverbal memory codes in terms of sequential and nonsequential processing properties.

The results of the present study are also relevant to this theoretical approach. Even though the verbal labels were readily available for the sounds—indeed labeling was required by the nature of the recall procedure—it seems reasonable to assume that verbal as opposed to nonverbal processing played a relatively greater role in the coding of words, and thus facilitated their serial recall. Given that this is so, it follows that verbal processing is more efficient when items are to be recalled in the correct serial order. As noted above, the same conclusions in support of Paivio's theory were reached by Paivio and Csapo (1969). However, it is not clear from their results whether the facilitative effect of verbal coding of material in sequential recall is attributable to verbal processes per se or to some general property related to the auditory sensory modality, to which the verbal code is functionally linked (Paivio, 1971). The results of the present experiment agree with those of Warren et al. (1969) in suggesting that verbal processing alone is the effective variable, since in both cases serial recall of auditory stimuli was enhanced by the relative predominance of the verbal memory code.

Finally, it is interesting to note that the nature of the Stimulus Class × Recall Task interaction observed with auditory material is not the same as that found with verbal and nonverbal (visual) material presented at a rate that permits verbal labeling of both types of stimuli to occur (Paivio & Csapo, 1969). Unlike the present results (Fig. 1), where words and sounds differed only in serial recall, examination of Fig. 2 of the Paivio and Csapo study shows that at the slow presentation rate, recall of pictures was higher than recall of the verbal labels in nonsequential tasks, while the two did not differ in sequential memory. The overall pattern of results from the two experiments suggests that pictures might be superior to sounds in both types of tasks. More specifically, recall of nonverbal visual material is facilitated in nonsequential tasks, while recall of nonverbal auditory material suffers in sequential tasks, relative to the recall of verbal material alone. Any speculation concerning the reason for this apparent difference in retention of nonverbal stimuli in the two modalities would be premature at this stage, but the findings do suggest a possible functional distinction between visual and auditory imaginal processes.

REFERENCES

Miller, J. D., & Tanis, D. C. Recognition memory for common sounds. *Psychonomic Science*, 1971, **23**, 307–308.
Paivio, A. *Imagery and verbal processes.* New York: Holt, Rinehart & Winston, 1971.
Paivio, A., & Csapo, K. Concrete image and verbal memory codes. *Journal of Experimental Psychology*, 1969, **80**, 279–285.
Sampson, J. R. Free recall of verbal and nonverbal stimuli. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 215–221.
Warren, R. M., Obusek, C. J., Farmer, R. M., & Warren, R. P. Auditory sequence: Confusion of patterns other than speech or music. *Science*, 1969, **164**, 586–587.

(Received March 22, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 344–346

INDEPENDENCE OF VERBAL AND VISUAL CODES OF THE SAME STIMULI [1]

HARRY P. BAHRICK [2] AND PHYLLIS BAHRICK

Ohio Wesleyan University

This study determines the relation between accuracy of visual recognition of previously exposed object drawings and probability of correct verbal recognition of the object names on a multiple-choice test. The results indicate that performance on the visual and verbal tests is uncorrelated. It is concluded that organizational class characteristics of visual storage do mediate verbal recognition, but that retention of these class characteristics is uncorrelated with retention of the item-specific aspects of visual storage needed for accurate visual recognition.

Information presented in visual form is often recoded verbally, and vice versa, and one code may serve to improve retention of the other. Thus, Bower (1970) has shown that visual imagery instructions during learning can improve recall performance of verbal material. Little is known, however, regarding the nature of the interrelations between visual and verbal codes during acquisition and retention. For example, the two codes may be maintained in storage independently, with one acting as a retrieval cue for the other; or the verbal code may be partially or completely lost and verbal retention depend upon recoding of certain aspects of visual storage at the time of recall.

Visual and verbal memory have frequently been compared, but the data are rarely based upon retention of the same stimuli by the same *S*s. In an earlier investigation (Bahrick & Boucher, 1968), probability of free recall of the names of object drawings was found to be independent of the accuracy of subsequent visual recognition of these same drawings. The *S*s were as likely to recall the object name of drawings they misidentified by committing large visual recognition errors as of drawings they subsequently correctly identified on a visual recognition test. Despite this lack of correlation of verbal and visual retention performance, it was concluded that *S*s used recoded visual storage to aid verbal recall. This interpretation was based on the assumption that recall of a verbal label for visual storage depends only upon the preservation of certain aspects of the visual code and that preservation of these critical aspects may be uncorrelated with the preservation of other aspects necessary for accurate visual recognition. Thus, *S* may be able to recall that he saw a drawing of a cup by retaining a general image of a cup, but retention of this general form may be uncorrelated with retention of other visual details such as the design on the cup or the shape of the handle.

An alternative interpretation of the observed independence of the two indicants of retention is based on the fact that the verbal test involved recall while the visual test involved recognition. The relation between recall and recognition is somewhat controversial (Tulving & Thomson, 1971), but there is much evidence that recall is greatly influenced by organizational variables which determine access to the desired material, while recognition performance is relatively unaffected by organization, since access to the material is provided through the recognition test (Bahrick, 1971; Kintsch, 1970). The observed independence of visual and verbal retention may thus reflect independence of the accessibility of the visual image from its accuracy. The visual image must be accessible in order to act as a mediator for verbal recall. If the accessibility is determined by factors uncorrelated with the accuracy of the image, then the accuracy of the image will be a poor predictor of verbal recall probability. To test this alternative explanation of the observed independence between verbal and visual performance in the present investigation, verbal recognition tests were substituted for the verbal recall tests.

Method.—The 16 previously used drawings of common objects (Bahrick & Boucher, 1968) were presented on an oversized memory drum at a 2-sec. rate to 40 undergraduate volunteer *S*s. Training conditions were the same as those described in the earlier study. The *S*s were alternately assigned to four groups of 10 *S*s arranged in accordance with a 2×2 design. As in the previous investigation, the independent variables in the design were the number of exposure trials of the list (1 or 9) and the verbal encoding instructions (free or instructed recoding). The four groups are designated respectively as 1 TR-IR, 1 TR-FR, 9 TR-IR and 9 TR-FR. The *S*s who served in 1 TR groups were shown the drawings in one of three random sequences. The *S*s who served in 9 TR groups were given a 20-sec. intertrial interval, and the three random sequences of drawings were alternated on successive trials. Free recoding (FR) groups were instructed only to observe the series of drawings carefully as they would be required to recognize them later. The instructed recoding (IR) groups were told in addition to call out the name of each object as the drawing appeared in the window. Two weeks after original exposure, all *S*s were given verbal recognition tests, followed immediately by visual recognition tests. The verbal tests consisted of 16 multiple-choice items presented

[1] This investigation was supported by Public Health Service Research Grant HD00926-09 from the National Institute of Child Health and Human Development.

[2] Requests for reprints should be sent to Harry P. Bahrick, Department of Psychology, Ohio Wesleyan University, Delaware, Ohio 43015.

on individual slips of paper. Each item consisted of five typewritten object names, one of which was the name of one of the object drawings exposed during training. The four wrong alternatives referred to object names, which had no strong association with, or obvious similarity to, any of the drawings that had been shown. The *Ss* were allowed 8 sec. to identify the correct alternative and were instructed to guess if necessary, making the procedure forced choice. They were then administered the next test item related to another drawing. The 16 test items were administered to each *S* in a different random sequence.

The visual recognition tests were identical to those described in the previous investigation. There were 16 test items and each item consisted of 11 drawings. One of these was the originally exposed drawings, while the 10 wrong alternatives were drawings of the same type of objects scaled to resemble the original drawings in 5 degrees, with 2 drawings representing each of five levels of similarity. The *Ss* were allowed 8 sec. to identify the correct drawing and were instructed to guess if necessary. The similarity rating scale and the testing procedure have been described previously (Bahrick, Clark, & Bahrick, 1967). In the earlier study, performance of control *Ss* who received only the visual test indicated that administration of a verbal test had little or no effect upon subsequent visual recognition performance. Comparable controls were omitted in the present investigation.

Results and interpretation.—Table 1 shows performance of the four groups on the visual and verbal recognition tests. Visual test performance was scored by categorizing responses as correct, or as first-, second-, third-, fourth-, or fifth-degree errors. As in the previous investigation, monotonic gradients of generalization can generally be plotted from the data of each group, and the slope of the gradient reflects the degree of training.

Verbal recognition performance is scored in terms of the total number of correct responses out of the 160 test items administered to each group. Analysis of variance of these data indicates the expected significant effects due to the degree of training, $F\ (1, 36) = 7.69$, $p < .01$. Instructed verbal encoding during training has very little effect, $F\ (1, 36) = 2.03$, $p > .05$, and the interaction of encoding instructions with the degree of training is not significant, $F\ (1, 36)$, $p < 1.00$. In the earlier study, instructed verbal encoding facilitated verbal recall only for the groups tested immediately after training, but not for the groups tested 2 wk. after training. This interaction led to the conclusion that verbal encoding was an important determinant of immediate verbal recall, but that 2 wk. after training verbal recall was based mostly upon recoded visual storage. This conclusion is in accord with other findings (Nickerson, 1968; Shepard, 1967) and is supported by the present data. The fact that 2 wk. after training verbal recognition performance is still very high, but shows no significant effect due to verbal encoding instructions, suggests strongly that the verbal recognition performance at that time is based more upon recoded visual storage than upon direct verbal storage.

TABLE 1

FREQUENCY DISTRIBUTION FOR RESPONSES ON THE VISUAL RECOGNITION TEST WITH ASSOCIATED VERBAL RECOGNITION FREQUENCY

Group	Similarity scale distance of testing stimuli from training stimuli						Verbal recognition frequency
	0	1	2	3	4	5	
1 TR-IR	41	43	27	20	14	15	94
1 TR-FR	32	31	37	24	23	13	81
9 TR-IR	49	41	32	21	14	3	132
9 TR-FR	50	46	30	18	8	8	120

The most important results of the present investigation concern the interrelations between verbal and visual recognition performance for the same drawings. This relation is shown in Table 2 in the form of conditional probabilities of verbal recognition given a certain level of accuracy of visual recognition of the same stimuli. It is apparent that the conditional probabilities vary systematically between rows, but not within the rows of the table. This means that verbal recognition probability is a function of the degree of training (as previously established) but is independent of the accuracy of visual recognition of the same stimuli.

The visual and verbal tests differ in several ways. The number of alternatives on each item of the test is not the same, and the visual test provides foils of scaled similarity to the correct choice, permitting multicategory scaling of error magnitude, while the verbal test can be scored only dichotomously since the similarity of foils to the correct alternatives is not determined. The effects of differential sensitivity of indicants upon the measurement of learning and retention have been discussed elsewhere (Bahrick, 1964, 1965) and the critical requirement for the determination of interrelations among indicants of learning is that both indicants must reflect acquisition changes during comparable stages of practice. Despite the differences between the verbal and visual tests, this requirement is clearly met. After a single exposure trial on the visual test, the correct alternatives are chosen nearly three times as frequently as the least similar foils, indicating great sensitivity to learning at this stage. The *Ss* trained for nine exposure trials chose the correct alternatives

TABLE 2

CONDITIONAL PROBABILITY OF VERBAL RECOGNITION AS A FUNCTION OF DEGREE OF ACCURACY OF VISUAL RECOGNITION FOR THE FOUR GROUPS

Group	Visual recognition error magnitude					
	0	1	2	3	4	5
1 TR-IR	.63	.51	.63	.45	.71	.60
1 TR-FR	.44	.55	.59	.58	.43	.31
9 TR-IR	.86	.76	.91	.71	.93	.67
9 TR-FR	.74	.78	.73	.67	.75	.87
$\bar{X}$	.67	.65	.71	.60	.71	.61

nearly 10 times as often as the least similar foils, indicating continued sensitivity of the test to the effects of training during this period.

The chance base line for the five alternative verbal test items is .20, i.e., a frequency of 32 correct responses out of the 160 responses made by each group of 10 *S*s. Actual performance is at approximately three times the chance level after one exposure trial, and the test continues to be sensitive to training during the later trials. Thus the visual and the verbal tests both reflect acquisition during comparable stages of training, but item-specific acquisition on the two tests is uncorrelated. Verbal recognition is equally probable for drawings correctly recognized and for drawings on which large recognition errors are committed. The previous investigation showed that verbal recall probability is independent of the accuracy of visual recognition, and the present data indicate that this is not due to independence of the accuracy of the visual code from its accessibility. Rather, the accuracy of the visual code is uncorrelated with the recognition probability of the corresponding verbal label. In light of the earlier conclusion that verbal recognition 2 wk. after training does depend upon recoding of visual storage, it follows that the mediating function of the visual code is independent of its fidelity or accuracy as measured by the visual recognition tests. Thus the mediation effectiveness of visual storage is unrelated to accuracy or completeness in relation to the stimulus. It is concluded that organizational or class characteristics of visual memory traces are retained independently of more specific item characteristics, and that the former, but not the latter, determine the effectiveness in the mediation of long-term verbal recognition or verbal recall.

REFERENCES

Bahrick, H. P. Retention curves: Facts or artifacts? *Psychological Bulletin*, 1964, **61**, 188–194.

Bahrick, B. P. The ebb of retention. *Psychological Review*, 1965, **72**, 60–73.

Bahrick, H. P. Accessibility and availability of retrieval cues in the retention of a categorized list. *Journal of Experimental Psychology*, 1971, **89**, 117–125.

Bahrick, H. P., & Boucher, B. Retention of visual and verbal codes of the same stimuli. *Journal of Experimental Psychology*, 1968, **78**, 417–442.

Bahrick, H. P., Clark, S., & Bahrick, P. Generalization gradients as indicants of learning and retention of a recognition task. *Journal of Experimental Psychology*, 1967, **75**, 464–471.

Bower, G. H. Mental imagery and associative learning. In Lee Gregg (Ed.), *Cognition in learning and memory*. New York: Wiley, 1970.

Kintsch, W. Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

Nickerson, R. S. A note on long-term recognition memory for pictorial material. *Psychonomic Science*, 1968, **11**, 58.

Shepard, R. N. Recognition memory for words, sentences, pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 156–163.

Tulving, E., & Thomson, D. M. Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, 1971, **87**, 116–124.

(Received March 25, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 347–350

MODES OF REPRESENTATION AND PROBLEM SOLVING: WELL EVOLVED IS HALF SOLVED

STEVEN H. SCHWARTZ[1]

Wayne State University

Thirty adult *S*s attempted to solve six "who-done-it" type deductive reasoning problems which varied in number of relevant dimensions, logical connective employed, and affirmative versus negative statement of information. Written protocols were classified into five modes of representation, with one—the matrix—clearly superior to other forms for affirmatively stated problems. Affirmatively stated problems were solved more easily than negative problems, with other task variables having little effect. The effect on solution rates of the task variables seemed dependent on their influencing the modes of representation *S*s employed. Possible reasons for the effectiveness of the matrix mode of representation were mentioned.

This research was designed to explore the relationships between how information in problem-solving tasks is represented or organized by adults and their subsequent performance. The effect that representation may have on cognitive activities has recently been stressed by theorists in perception, verbal learning, artificial intelligence, and problem solving (Mandler, 1967; Posner, 1969; Simon & Newell, 1971). Such a concern is not new to psychology; it represents a re-emergence of the Gestalt emphasis on "restructuring the field," "reorganizing the task," or "seeing the problem a new way" (Dunker, 1945; Wertheimer, 1956). A reconsideration of these problems using some of the constructs derived from recent information-processing approaches to learning and cognition (Neisser, 1967; Norman, 1970; Reitman, 1969) now appears promising.

To assess the effects of representation, tasks were designed which required *S* to make overt his organization of information. "Who-done-it" type deductive reasoning problems generate such written protocols for nearly all *S*s, since the memory load is normally too great to solve these problems "in your head."

Conjunctive positive problem
(3 dimensions—5 values):

Five men were in a hospital. Each one is suffering from a different disease.

1. The man with asthma is in Room 101.
2. Mr. Alex has cancer.
3. Mr. Osborn is in Room 105.
4. Mr. Wilson has TB.
5. The man with mononucleosis is in Room 104.
6. Mr. Thomas is in Room 101.
7. Mr. Wilson is in Room 102.
8. One of the men has epilepsy.
9. One of the patients is in Room 103.

WHAT DISEASE DOES MR. YOUNG HAVE?

Conjunctive negative problem
(4 dimensions—5 values):

Five different men of five nationalities live in five different colored houses. Each one has a pet and a favorite drink.

1. The hyena's owner doesn't live in the white, yellow, or green house.
2. Neither the Japanese, the Indian nor the Englishman lives in the green house.
3. Neither the American nor Canadian owns a zebra.
4. The tea drinker doesn't live in the blue house and doesn't own a turtle, hyena, or ox.
5. Neither the Japanese nor the Englishman owns a hyena.
6. The beer drinker isn't English, doesn't live in the red house, and doesn't own an ox.
7. The zebra's owner doesn't live in the yellow or red house and doesn't drink milk.
8. The coffee drinker doesn't own a zebra or live in the yellow house.
9. The Japanese doesn't live in the red, blue, green, or yellow house.
10. The American doesn't live in the red, blue, green, or yellow house.
11. One of the men drinks whiskey.

WHAT DOES THE OWNER OF THE HORSE DRINK?

Problems differing on a number of task variables known to have considerable effect on traditional concept-attainment tasks (Bourne, Ekstrand, & Dominowski, 1971) were constructed in order to determine their effects on *S*s' representation of information.

Method.—Thirty-eight *S*s were randomly drawn from the *S* pool at a large urban university and tested individually in sessions of approximately 1½ to 2 hr. (Eight of these *S*s were eliminated from the study on the basis of failing to solve or make any progress on a simple test problem in 15 min.)

[1] Requests for reprints should be sent to Steven H. Schwartz, Department of Psychology, 746 MacKenzie Hall, Wayne State University, Detroit, Michigan 48202.

TABLE 1

PERCENTAGE OF SUBJECTS SOLVING UNDER EACH CONDITION

Logical relation	No. of dimensions				Total
	Three		Four		
	Affirmative	Negative	Affirmative	Negative	
Conjunctive	80%	20%	93%	13%	52%
Disjunctive	87%	27%	67%	13%	48%
Conditional	73%	40%	60%	33%	52%
Total	80%	29%	73%	22%	51%

Note.—N = 15 per cell.

The study employed a mixed design with number of dimensions as a between-*Ss* factor. Twelve "who-done-it" type deductive problems were constructed in terms of the following task considerations: (*a*) In half of the problems the information was presented in an affirmative fashion (i.e., The redhead bought the miniskirt), while the other six problems contained almost exclusively negative information (i.e., The man in Room 101 does not have cancer, emphysema, or mononucleosis). (*b*) The information in each set of four problems used either conjunctive (Ed had coffee *and* pie), disjunctive (Dave either wears a blue *or* red tie), or conditional (*If* Atlanta got the pitcher, *then* Montreal got the shortstop.) connectives. (*c*) Finally, half of the *Ss* worked on problems with five values on *three* dimensions (e.g., name of girl, article of clothing, and color), while the other *Ss*' problems contained five values on *four* dimensions (e.g., name of suspect, location of crime, murder weapon, and motive). There was sufficient information in each problem for determination of a unique solution.

Each *S* was encouraged to show all work and had a maximum of 18 min. to attempt to solve each of the six problems. The order of problems was randomized. If *S* solved in less than the maximum allowed time, he went on to the next problem. After six problems, *S* took a brief multiple-choice logic test which was designed to indicate the extent to which he could draw valid implications from *single* affirmative or negative conjunctive, disjunctive, and conditional statements.

A coding scheme developed from pilot data classified *Ss*' protocols into the following five modes of representation: (*a*) a *matrix* representation indicating an explicit Row × Column organization of dimensions and values; (*b*) an *informal grouping* representation where information was circled, placed in close proximity, connected by a hyphen, etc.; (*c*) a *graphic* representation covering items connected by extensive series of lines often resembling trees; (*d*) a *sentence* representation consisting of simply re-writing some of the sentences in a different order; (*e*) a *miscellaneous* category covering those protocols that did not fit into any of the above categories—often little at all was written down.

Results.—Success of performance in terms of percentage of *Ss* solving each problem is indicated in Table 1. (Time to solution yielded similar results and therefore is not discussed). The analysis of variance yields a large main effect, $F(1, 28) = 92.73$, $p < .001$, in terms of whether the information in a problem is presented in affirmative or negative fashion. Affirmatively presented problems are easier to solve in all conditions studied.

A significant interaction was also found between type of logical connective and affirmative versus negative forms, $F(2, 56) = 4.60$, $p < .02$. Although negative problems are more difficult under all conditions, the magnitude of the difference is smaller in the conditional as compared to conjunctive or disjunctive problems. (There are about half as many solvers in negative as in affirmative conditional problems, compared to about one-quarter as many solvers in negative vs. affirmative conjunctive, or disjunctive problems.) All other main effects and interactions were insignificant ($ps > .30$).

There were no significant order effects, $F < 1.0$, nor any significant relation between problem-solving performance and performance on the logical inference test ($r = .25$, $p > .05$), indicating that more than the ability to draw correct inferences from single statements is required to solve these problems.

Analysis of the protocols in terms of *Ss*' modes of representation as classified by the scheme previously described (interrater correlation of .88 between two independent raters) is presented in Table 2. Before examining the table, note that *Ss* were rather consistent in using a dominant mode of representation for all problems; 80% used the same representation on at least four of the six problems they attempted.

A number of findings in Table 2 are of particular interest.

1. Considering all 180 problems one mode of representation, the matrix, was superior to all others in terms of success rate achieved (74% solved compared to about 50% or less for other representations $\chi^2(4) = 14.12$, $p < .01$). Further, the success rate for those problems with little or no representation (Category E) was significantly lower $\chi^2(1) = 4.99$, $p < .05$, than even the less successful modes (B, C, and D).

2. The *Ss* employed the matrix representation almost twice as often on affirmative as compared to negative problems ($p < .01$). No such differences were found for any other form of representation, ($p > .50$ in all other cases).

3. Neither the distribution of usage nor solution rates for various modes of representation vary with type of logical connectives or number of dimensions used in problems (all $ps > .30$).

Discussion.—The comparative solution rates for the different types of problems presented in Table 1 should be interpreted with caution because of the problem of precisely equating the information in each problem. In one sense, the information presented in each problem is equivalent and sufficient. Unique value assignments can be deduced for all dimensions in every problem. Likewise, each problem contained two to three redundant or irrelevant statements. However, information directly presented in one problem, i.e., "Bob had cake for

TABLE 2

Proportion of Times Each Mode of Representation Was Used and Corresponding Solution Rates on Various Types of Problems

Mode of representation	Affirmative (90 Problems)		Negative (90 Problems)		3 Dimensions (90 Problems)		4 Dimensions (90 Problems)		Conjunctive (60 Problems)		Disjunctive (60 Problems)		Conditional (60 Problems)		Overall (180 Problems)	
	Use	Solution	Use	Solution	Use	Solution	Use	Solution	Use	Solution	Use	Solution	Use	Solution	Use	Solution
A: Matrix	.33	.97	.18	.37	.22	.84	.30	.67	.25	.73	.28	.71	.23	.79	.26	.74
B: Informal grouping	.49	.70	.56	.21	.50	.47	.54	.43	.50	.50	.52	.39	.55	.45	.52	.45
C: Graph	.08	.71	.03	.25	.10	.67	.02	.00	.05	.67	.07	.50	.07	.50	.06	.55
D: Sentence	.03	.50	.03	.33	.04	.50	.01	.00	.02	.00	.03	.50	.03	.50	.03	.40
E: Miscellaneous	.07	.57	.19	.12	.14	.12	.12	.18	.18	.27	.10	.17	.12	.29	.13	.25

Note.—N = 180 total problems.

dessert," may be inherent in the combination of two or more statements in an alternative problem, i.e., "Bob did not eat fruit or jello. Bob did not eat ice cream or pie. The five desserts available are fruit, jello, ice cream, pie, and cake." In both cases, we can correctly conclude Bob had cake, but the latter presentation required a logical combination of three statements. Hence, one could claim the information *directly* presented is less in the latter case. Since our primary purpose was to explore relationships between representation and performance on a variety of problems, such a statement-by-statement equating of information was not undertaken. (A study evaluating performance under such conditions is in progress).

With these cautions in mind, note that unlike typical findings on concept-attainment tasks neither number of relevant dimensions nor type of logical relation had any significant effect on number of *S*s attaining solution. However, as expected, problems with information presented affirmatively were significantly easier than negative problems. The reason for these contrary findings is suggested in the use of various modes of representation as indicated in Table 2. Note the only significant difference in the distribution of modes of representation used occurs between the affirmative and negative problems with almost twice as much use of the matrix representation in the affirmative problems. The other task variables (number of relevant dimensions and logical relation) had little influence on mode of representation employed. Hence it appears that the effects of various task variables on solution rate depend on whether or not these variables influence the modes of representation *S*s will employ. Only when they do so can one expect differential solution rates.

The study lends support to the notion of the significance of mode of representation of information on problem-solving performance. Of the five types of representations commonly used, the matrix representation was significantly superior to other types of representation yielding about 75% solutions compared to 50% for other types and 15% for little or no representations. That different representations may be of differential utility in different problems is suggested by the fact that of 30 attempts to solve positive problems by matrix representation, 28 were successful compared to only 6 successful attempts in 16 negative problems. Hence, not only were *S*s less inclined to transfer the sentences containing negative information into a matrix format, but when they did so, they typically did not attain solutions.

Further analysis of the 10 unsuccessful attempts to solve negatively stated problems indicates that in all but one case there were errors in representing the original information in a matrix format. Since the given information was not accurately represented to begin with, subsequent cognitive operations were in a sense doomed to failure. In six of the seven cases where *S* was able to correctly place the negative information into a matrix format, correct solutions were obtained. Success on deductive reasoning problems such as those presented here seems to be due in large measure to the ability to accurately represent the given information in a form suitable for carrying out further manipulations.

The more fundamental question of why a matrix representation was so effective for these types of problems is open to speculation. To suggest several of many possibilities, this form of representation clearly defines needed information (empty cells), suggests fruitful orders of operation (work on rows or columns with minimal missing information), and provides easy consistency checks for part solutions (attribute cannot occur twice in same row or column).

Studies are currently in progress which attempt to delineate between these and a number of other alternatives, as well as investigate other types of tasks for which different representations appear more optimal.

REFERENCES

BOURNE, L. E., JR., EKSTRAND, B. R., & DOMINOWSKI, R. L. *The psychology of thinking.* Englewood Cliffs: Prentice-Hall, 1971.

DUNKER, L. On problem-solving. *Psychological Monographs,* 1945, **58**, (5, Whole No. 270).

MANDLER, G. Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 1. New York: Academic Press, 1967.

NORMAN, D. A. *Models of human memory.* New York: Academic Press, 1970.

POSNER, M. I. Perception and cognition as processing models. In J. F. Voss (Ed.), *Approaches to thought.* Columbus: Charles E. Merrill, 1969.

REITMAN, W. Information-processing models, computer simulation and the psychology of thinking. In J. F. Voss (Ed.), *Approaches to thought.* Columbus: Charles E. Merrill, 1969.

SIMON, H., & NEWELL, A. Human problem solving. The state of the theory in 1970. *American Psychologist,* 1971, **26**, 145–159.

WERTHEIMER, M. *Productive thinking.* New York: Harper, 1956.

(Received June 11, 1971)

Journal of Experimental Psychology
1971, Vol. 91, No. 2, 350–353

HIGH-SPEED VISUAL SCANNING OF WORDS AND NONWORDS[1]

NEIL NOVIK[2] AND LEONARD KATZ

University of Connecticut

Letter scanning rates were compared for words and nonwords. In Exp. I, mixed stimulus lists for words and nonwords were presented. In Exp. II, *S*s received either words or nonwords. In both experiments, *S*s scanned words faster than nonwords. The results suggested sequential, exhaustive scanning.

Recently, Kreuger (1970) demonstrated that adult *S*s can search for a single letter faster through an array of real words than through an array of nonwords. His *S*s were presented with a target letter followed by an array of 25 six-letter words or nonwords arranged in five rows of five items each. The *S*s were required to respond "yes" or "no," depending on whether the target letter appeared in the array. By measuring the reaction time (RT) as a function of the row in which the target letter appeared, Kreuger estimated that a single word was processed about 100 msec. faster than a nonword.

These results indicated that *S*s were encoding the stimuli at least to the extent that interletter dependencies were utilized to aid in the search of a target. However, Kreuger's design does not allow for a more detailed comparison of the visual processing of words and nonwords. In an experiment by Atkinson, Holmgren, and Juola (1969), *S*s were presented with a target letter, followed by a horizontal display ranging from one to five letters. The *S*s were required to report on the presence or absence of the target letter in the display. Atkinson et al. found linear increases in RT with increasing display size and no slope differences between positive and negative responses. These results suggested that *S*s were using a high-speed, serial, exhaustive scan (cf. Sternberg, 1969). Also, with this method, which produces RTs in the 500-msec. to 1,000-msec. range, a more precise measure of visual attention time per letter can be obtained than in the Kreuger (1970) experiment because large saccadic eye movements are few or absent.

The present studies were designed to compare visual scanning of both words and nonwords, using a modification of the technique of Atkinson et al. (1969). This allowed for a more accurate estimation of the processing time of the displays as well as a more complete analysis of differences between words and nonwords.

Experiment I.—Eighteen right-handed male undergraduate students at the University of Connecticut served as volunteer *S*s. An additional 12 *S*s were run, but had to be replaced due to subcriterion performance (described below).

The stimuli were typed on an ASR-33 teletype with a mask placed so that *S* could not see the line as it was being typed, but only after the completed line had been advanced to a window in the mask. The window exposed only one line of type at a time. A Data General computer, model Nova, fed all the stimuli to the teletype and recorded responses and RTs.

On each trial, *S* saw two successive stimulus lines; the first was always a single letter (the key letter) and the second was a string of three, four, or five letters (the display), which was typed below and two spaces to the right of the key letter. The task was to scan each display for the presence or absence of the immediately preceding key letter.

The two display types were words and nonwords. For the test trials, 20 common, one-syllable words of each length were chosen with the restriction that no letter appeared in a given word more than once. The

[1] This study was supported in part by Grant HD-03932 from the National Institute of Child Health and Human Development to the second author and to D. A. Wicklund, and in part by National Institutes of Health Training Grant HD-00084-07 from the United States Department of Health, Education, and Welfare to the first author. The authors would like to thank Leslie Waite for assistance in the collection of data and Shang-Wu Kuo for assistance in data analysis.

[2] Requests for reprints should be sent to Neil Novik, Department of Psychology, Box U-20, University of Connecticut, Storrs, Connecticut 06268.

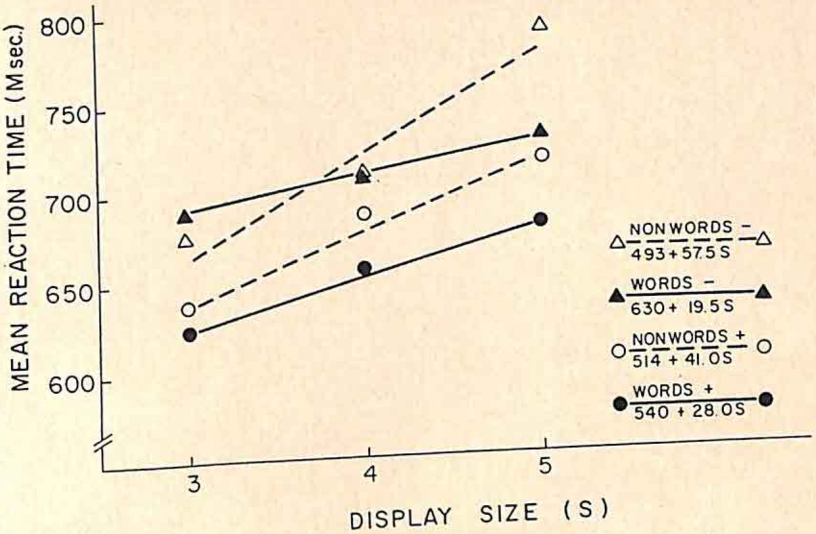

FIG. 1. Experiment I: Mean RT as a function of display size (in msec.).

nonwords were formed by scrambling the letters of each word so that the resulting nonwords would have lower interletter dependencies. As a measure of interletter dependency, the average bigram frequency for each display stimulus was tabulated on the basis of occurrence data provided by Underwood and Schulz (1960). The mean bigram frequencies for the three-, four-, and five-letter words were 1,233, 1,442, 1,551, respectively, while the mean frequencies for the nonwords were 719, 670, and 921, respectively.

The trials were blocked by display size only. Within each block, there were four conditions made up of the factorial combinations of word and nonword with positive and negative responses. There were 10 trials for each condition and these were randomly mixed within a block. For the positive trials, the key letters were quasibalanced for serial position within the displays. For the negative trials, the key letters were randomly chosen from the remaining letters of the alphabet.

Three blocks of practice trials were constructed with different stimuli in the same manner as described above. Each *S* received all the practice trials first. The order of block presentation was counterbalanced.

The instructions consisted of a description of the task and apparatus; *S*s were also told to respond as quickly and as accurately as possible.

For each trial, the key letter was exposed for 1 sec. and was immediately followed by the display. Depending on whether the key letter appeared in the display, *S* was required to press one of two buttons on the teletype, designated for "yes" and "no" responses, with his right and left hand, respectively. The display remained on for a maximum of 3.5 sec. If *S* had not responded by that time, an error was recorded and the next key letter was presented. Each display was immediately terminated upon *S*'s response, and 1 sec. separated the response and the onset of the next key letter.

There was about a 3-min. interval between the presentation of each block of 40 trials, during which time *E* collected the data of the previous block and fed the stimulus tape of the next block into the computer.

Failure criterion was a median RT of above 900 msec. for any test list, or more than 7% errors on all three test lists.

A repeated-measures analysis of variance was performed on the median RTs for all correct responses on each test list. The three factors were type of display (word vs. nonword), size of display (three, four, and five letters), and response mode (yes vs. no).

Means across *S*s for the median RTs were computed for each condition. Summary data are presented in Fig. 1, along with least-squares linear fits for each curve. All main effects attained statistical significance. There was an increase in RT due to increasing display size, $F\ (2, 34) = 29.20$, $p < .001$. The linearity of the data suggest serial scanning for both words and nonwords. Faster RT was also found for words than for nonwords, $F\ (1, 17) = 16.77$, $p < .001$, and for yes responses than for no responses $F\ (1, 17) = 41.77$, $p < .001$. There was no first-order interaction involving response mode; this absence, together with inspection of the data suggest that scanning was exhaustive. The most important finding was a significant interaction of size and type of display, $F\ (2, 34) = 12.93$, $p < .001$. Averaging across response mode, the least-squares equations for RT in milliseconds for words and nonwords were $586 + 23.5S$ and $505 + 49.0S$, respectively, indicating that the scanning rate for words was more than twice as fast as that for nonwords.

Finally, there was a tendency toward a faster

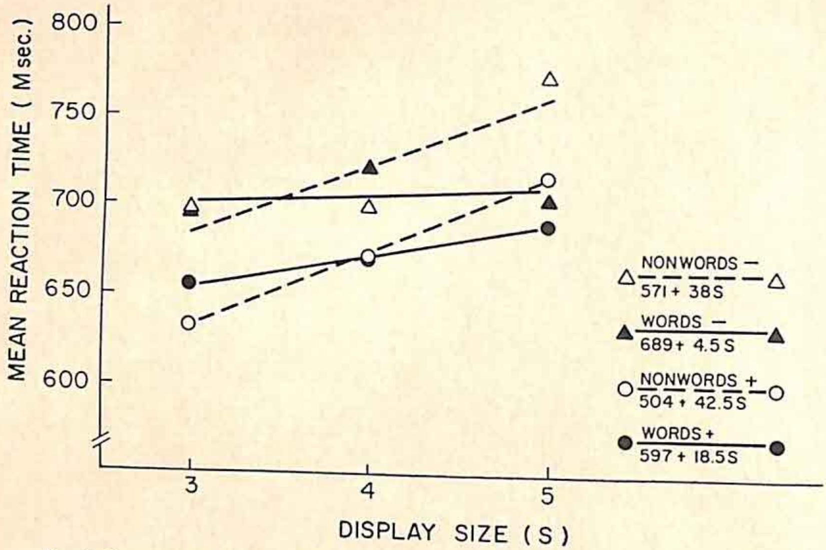

FIG. 2. Experiment II: Mean RT as a function of display size (in msec.)

scanning rate for negative responses with words and a slower scanning rate for negative responses with nonwords compared to their respective positive response scan rates. The Display Size × Type of Display × Response Mode interaction attained significance, F (2, 34) = 3.29, $p < .05$.

The results of Exp. I suggested that *S*s were able to utilize the higher intraword redundancy of real words in high-speed scanning. They may have done so by adopting a set for words, that is, by performing in a "word" mode; the scanning of nonwords would not be facilitated in this mode and may have been made even slower than a condition in which nonwords alone were presented. Therefore, a second experiment was run in order to assess the possibility that the mechanism which utilizes intraword redundancy is under *S*'s control. Specifically, if the word-nonword factor was made a between-*S*s factor, each *S* would be free to adopt whatever strategy maximized his scanning rate in his condition.

Experiment II.—Twenty-four right-handed male undergraduate students served as *S*s. Five *S*s exceeded the failure criterion and were replaced.

Experiment II employed a mixed design, with display type as a between-*S*s factor, while display size and response mode remained within-*S*s factors. Thus, all *S*s were tested either on words or on nonwords, eliminating the uncertainty of display type. The same stimuli used in Exp. I were again used in Exp. II, except that they were blocked by display type as well as by display size. Each test block, therefore, contained only 20 trials. There were 40 trials in each practice block. These trials consisted of the 20 words or nonwords used as practice in Exp. I plus an additional 20 items per block. Each *S*, therefore, received three 40-trial practice blocks followed by three 20-trial test blocks. All other procedures were the same as in Exp. I.

An analysis of variance was performed on the median RTs for correct responses on the test lists. The factors were identical to those in Exp. I, except that display type was a between-*S*s factor.

The only main effects to attain statistical significance were display size and response mode. As in Exp. I, RT was slower for negative responses than for positive responses, F (1, 22) = 17.53, $p < .001$, and also increased with increasing display size, F (2, 44) = 12.09, $p < .001$. Summary data for Exp. II are presented in Fig. 2. There was no main effect due to display type, but the Display Type × Size of Display interaction was significant, F (2, 44) = 4.35, $p < .025$. This replicated the finding of Exp. I that words were scanned faster than nonwords. No other interactions were significant.

Discussion.—These results support those obtained by Kreuger (1970) indicating that words are scanned faster than nonwords. This was evident in both experiments by significant interactions of size and type of display.

Although Atkinson et al. (1969) used nonwords as their display stimuli, their scanning rates were almost twice as fast as the scanning rates for nonwords found in the present study. This difference can probably be attributed to the greater number of trials experienced by *S*s in the former study. The lack of a significant first-order interaction with response mode in either of the present experiments, however, supports results obtained by Atkinson et al. suggesting that *S*s were scanning the displays exhaustively. Also, the large slopes for nonwords, at least, strongly suggest serial processing, although no information is provided about the size of the unit, e.g., letter, bigram, or other. However, while serial processing of words seemed to be indicated by the data of Exp. I, it was not quite as evident in the data of Exp. II. The slope for words

shown in Fig. 2 suggests that *S*s were scanning letters at the rate of about 90 per second, which is quite fast (compared to nonwords) for serial processing. It is possible that with additional practice, *S*s would have shown even faster scanning rates—approaching parallel processing—for word stimuli.

It is still unclear as to which aspects of words are essential in order for faster scanning to occur. Kreuger (1970) found that RT was slower for rare words than for common words and slower still for third order pseudowords. Since all of his displays consisted of two, six-letter items, there was still no way to compare actual scanning rate for the three types of displays.

Reicher (1969), using tachistoscopic presentation, found that recognition of a letter was best when the letter was embedded in a word than when it was embedded in a nonword or presented alone. More recently, Wheeler (1970) has tentatively suggested that the superior performance on words found in Reicher's study may be due to (*a*) the utilization of visual features from digrams or larger units, (*b*) greater efficiency in selecting letter features, and/or (*c*) the utilization of memory for word names in order to decrease the uncertainty about the target letter. The same processes are, of course, applicable to the scanning results of the present study. Research is being conducted in the present authors' laboratory to investigate the effects of interletter dependencies, word frequency, and pronounceability on scanning rates and strategies.

REFERENCES

Atkinson, R. C., Holmgren, J. E., & Juola, J. F. Processing time as influenced by the number of elements in a visual display. *Perception & Psychophysics*, 1969, **6**, 321–326.

Kreuger, L. E. Search time in a redundant visual display. *Journal of Experimental Psychology*, 1970, **83**, 391–399.

Reicher, G. M. Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 1970, **81**, 275–280.

Sternberg, S. Memory-scanning: Mental processes revealed by reaction time experiments. *American Scientist*, 1969, **57**, 421–457.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning*. Chicago: J. B. Lippincott, 1960.

Wheeler, D. D. Processes in word recognition. *Cognitive Psychology*, 1970, **1**, 59–85.

(Received May 3, 1971)

ERRATUM

In the article "The Total Time Hypothesis: A Reply to Stubin, Heurer, and Tatz" by B. R. Bugelski and M. L. McMahon in the September 1971 issue, the name Heurer in the title and in paragraph 1, line 1 should read Heimer, i.e., Stubin, Heimer, and Tatz (1970).

JOURNAL OF EXPERIMENTAL PSYCHOLOGY
MONOGRAPH

Vol. 91, No. 2, 355–364 December 1971

CATEGORIZATION NORMS FOR FIFTY REPRESENTATIVE INSTANCES[1]

ELIZABETH F. LOFTUS[2] AND RONALD W. SCHEFF

New School for Social Research

Two hundred *S*s were presented 50 single-word nouns and asked to provide three categories into which each noun belonged. For each stimulus, the frequency of occurrence of each of the 10 most frequent responses in each of the three successive positions is tabulated.

Investigations of conceptual organization and class membership have made extensive use of norms collected by asking *S*s to provide specific instances of common taxonomic categories. Two procedures for obtaining these norms have been used. Cohen, Bousfield, and Whitmarsh (1957) and Shapiro and Palermo (1970) presented category names and instructed *S*s to list the first four items that they thought of as representative members of each category. Battig and Montague (1969) read the name of the category to *S*s and gave them 30 sec. in which to write down as many members of that category as they could. In both procedures, the frequency of occurrence of each response to each category name was summed across *S*s, resulting in a frequency distribution for each category. This provides an index of the relatedness of a specific instance to the general category.

These norms have been invaluable to *E*s studying conceptual processes. They have been the major source employed for the construction of lists in studies of category clustering (e.g., Bousfield, Cohen, & Whitmarsh, 1958) and have facilitated research on semantic memory (e.g., Loftus & Freedman, 1970). Recently, Freedman and Loftus (1971) used the Battig and Montague (1969) norms and showed that the higher the normative frequency of the instance word as an associate to the category name, the faster *S*s can retrieve that instance from memory.

Since a considerable amount of current research involves procedures in which *S*s are asked to categorize words, it would be helpful to have normative data directly relevant to this task. Although the available norms are very useful, they are limited in that they indicate the strength of relationship between a category name and a member of that category, but provide no measure of the reverse relationship. For example, 37% of *S*s who were asked to name instances of the category SNAKES listed COPPERHEAD (Battig & Montague, 1969), while only 1% of *S*s who were asked for members of the category REPTILES listed COPPERHEAD (Shapiro & Palermo, 1970). This raises the question of what responses *S*s would give to the instance COPPERHEAD if they were asked to think of categories to which it belonged. As we shall see later, this reverse procedure produces quite different results—89% of *S*s gave SNAKE and 26% gave REPTILE. Thus, even though COPPERHEAD is a low-frequency response to REPTILES, REPTILE is a relatively strong associate to COPPERHEAD. The present study was designed to provide a set of norms of the frequency with which superordinate categories are given to particular instances.

[1] The authors thank Enid Natkins for her appreciable contributions to this research.

[2] Requests for reprints should be sent to Elizabeth F. Loftus, The Graduate Faculty, New School for Social Research, 66 West 12th Street, New York, New York 10011.

Method

Two hundred students at the New School for for Social Research were presented with 50 instances and instructed to list three categories to which each instance belonged. A group-testing procedure involving a booklet technique was used, with groups ranging in size from 1 to 103 *S*s.

Fifty single-word nouns were selected so as to be representative members of a wide variety of categories. To minimize contextual effects, three random orders were prepared with the restriction that similarities between adjacent items were minimized within each order. In addition, when *S* gives more than one response to a particular stimulus, there is always the problem that later responses may be elicited by a combination of earlier responses and the stimulus word. As a way of minimizing this problem, we required *S*s to give only one response at a time, and arranged sequences so that a considerable number (50, on average) of other stimuli occurred between two appearances of the same stimulus. The words were presented in three random orders, each *S* receiving all three orders in one of the six possible permutations. Stimuli were presented on mimeographed sheets containing all 50 words arranged in two vertical columns. Each *S* received three sheets, stapled together to form a booklet. The *S*s were required to give one response to each word on each sheet. Thus, *S*s saw each stimulus three times, once on each of the three sheets, and gave three different responses to it.

The following instructions were given:

> Previous investigators have presented subjects with a category name and asked them to think of items that are representative members of the specified category. For example, when asked to name items that belong to the category "colors," subjects might say "red," "blue," or "green." When asked to name "countries," subjects might say "France," "Russia," or "Germany." In this experiment we are going to do just the opposite. You will be presented with a noun and we would like you to respond with a category to which it belongs. Give the first category you think of. If you see the noun a second time, try to respond with a different category name to which it belongs. If you see the noun a third time, respond with a category name that is different from your first two. Continue until you have done this for all nouns on all sheets of the booklet.

Note that no explicit examples of the required task were given because we felt that examples might bias *S*s responding. To illustrate, if we had used POODLE as our example, and suggested the category names DOG, ANIMAL, and LIVING THING as possible responses, we might have encouraged the use of the category LIVING THING where it otherwise might not have been used.

Results and Discussion

The data are shown in 50 separate tables. For each stimulus, we have reported the 10 most frequent responses. Only 10 responses are given for each stimulus because the frequencies are quite low for responses beyond this point. The frequency, out of 200, with which each of the 10 response words was given in each of the three successive positions is shown under the columns headed 1, 2, and 3. The sum of these three frequencies yields the number of *S*s who gave a particular response, and this total is listed in the last column on the right.

In tabulating the data, different forms (i.e., misspelled words, spelling variants, singular and plural responses) were combined. There were occasional illegible responses and, in some cases, *S*s wrote fewer than three words in response to a particular stimulus, or used a particular category name more than once. In the latter case, the second occurrence was ignored.

The order in which the instances are tabled is alphabetical. The normative responses to each instance are listed in order from most to the least frequent total number of responses.

At the conclusion of this experiment, many *S*s volunteered the information that they found the task difficult and at times frustrating. The reader can easily appreciate this difficulty by trying to think of three categories for the instances CEILING and DISH. The number of *S*s responding with "incorrect" categories may provide an indication of the difficulty of categorizing particular instances. For example, there were 26 *S*s (13%) who gave the response HIGH to the stimulus CEILING, and 20 *S*s (10%) gave the response SHAPE to the stimulus DISH. An inspection of the entire response protocols shows that this was not a matter of misunderstanding the instructions as these same *S*s were able to respond with INSECT to BUTTERFLY, with FRUIT to CANTALOUPE, and with DOG to COLLIE.

The instances CEILING and DISH are two examples of words which do not have a well-defined, universal category into which they are classified. The most common response to CEILING, PART OF BUILDING, was

given by 40% of the *S*s; the most common response to DISH, UTENSIL, was given by 46% of the *S*s. These provide a contrast with examples such as PEAR, SPARROW, and DAISY, to which 96% gave FRUIT, 97% gave BIRD, and 98% gave FLOWER, respectively. The finding that the stimuli varied in terms of the frequency with which the most common response was given mirrors a result found with the reverse procedure. When presented with the category ARTICLE OF FURNITURE, 100% of the *S*s said CHAIR; to the category name TYPE OF SHIP, the most common response was SAILBOAT, given by only 40% of the *S*s. (Battig & Montague, 1969). Thus, regardless of the procedure, some stimuli have very strong associates and others do not.

An interesting aspect of the results which was mentioned earlier is that even though a particular stimulus word is an uncommon response to a particular category name, the category name may be a relatively common response to the stimulus word. BUTTERFLY is a low-frequency response to INSECT, but INSECT is a very strong associate to BUTTERFLY; CHIMPANZEE is a very rare associate to ANIMAL, but ANIMAL is a high-frequency response to CHIMPANZEE, to name a few. Research on semantic memory and organizational processes must take these data into account.

TABLE 2

Response	Aluminum			
	1	2	3	Total
metal	113	23	1	137
alloy	13	12	5	30
material	3	13	11	27
element	7	11	5	23
light-weight material	1	11	9	21
can	5	5	9	19
type of wrapping paper	1	11	3	15
reflecting material	0	10	5	15
ore	6	3	2	11
mineral	4	3	2	9

TABLE 3

Response	Boat			
	1	2	3	Total
transportation (means of transportation)	50	13	3	66
water vehicle	23	23	7	53
type of ship	27	11	3	43
vehicle	25	13	5	43
floating object	4	25	13	42
vessel	9	13	1	23
yacht	8	9	2	19
sailing vessel	9	2	1	12
pleasure craft	3	7	0	10
craft	5	0	1	6

TABLE 1

Response	Adultery			
	1	2	3	Total
sin	39	30	5	74
crime	25	21	7	53
sex	24	5	11	40
marriage	9	15	1	25
act	3	11	7	21
fun	3	10	8	21
divorce (grounds for divorce)	4	11	5	20
sexual act (sexual behavior)	9	8	1	18
morals (morality)	7	3	3	13
behavior	6	3	3	12

TABLE 4

Response	Burglary			
	1	2	3	Total
crime	133	13	7	140
type of theft	15	27	7	49
act	15	11	19	45
act of stealing	5	23	1	29
felony	3	21	5	29
type of robbery	3	9	7	19
antisocial act	5	5	6	16
sin	3	8	1	12
jailable offense	0	2	7	9
evil	1	2	5	8

TABLE 5

Response	Butterfly 1	2	3	Total
insect	147	7	5	159
animal	29	23	5	57
flying (flying object)	1	25	17	43
winged (winged insect, winged object)	2	15	8	25
flying insect	5	9	5	19
bug	7	5	2	14
beautiful object (pretty object)	1	7	6	14
colorful (colorful object)	1	7	1	9
flying animal	1	5	2	8
living thing	1	2	5	8

TABLE 6

Response	Cantaloupe 1	2	3	Total
fruit	125	13	2	140
melon	35	33	1	69
food	17	23	17	57
round (round object)	3	23	9	35
orange (orange object)	1	17	11	29
dessert	7	6	5	18
seeded fruit	3	5	9	17
sweet (sweet food)	2	9	4	15
yellow (yellow object)	1	7	3	11
summer fruit (seasonal fruit)	1	3	5	9

TABLE 7

Response	Cathedral 1	2	3	Total
building	89	35	7	131
type of church	79	35	3	117
place of worship	13	34	11	58
religious building	12	11	3	26
religious symbol	9	10	4	23
architectural form	3	3	2	8
sanctuary	1	4	3	8
spired building	0	5	2	7
massive building	0	4	3	7
Gothic building	1	3	2	6

TABLE 8

Response	Ceiling 1	2	3	Total
part of building	61	13	5	79
part of room	42	5	3	50
top	9	18	13	40
roof	10	15	5	30
high	5	10	11	26
covering	14	9	1	24
upper limit	8	12	3	23
structural position	5	4	1	10
height	2	6	2	10
architectural unit	4	2	1	7

TABLE 9

Response	Champagne 1	2	3	Total
drink	88	17	9	114
carbonated beverage	2	46	21	69
alcoholic beverage	24	13	15	52
wine	21	17	7	45
liquor	23	11	1	35
liquid	7	13	11	31
beverage	23	3	2	28
intoxicant	8	7	5	20
celebration	1	8	6	15
refreshment	0	3	3	6

TABLE 10

Response	Chimpanzee 1	2	3	Total
animal	128	31	3	162
primate	30	17	8	55
monkey	15	31	9	55
mammal	10	33	10	53
ape	13	17	15	45
living thing	1	6	11	18
pet	1	7	9	17
vertebrate	1	5	7	13
experimental animal (test animal)	0	3	6	9
intelligent animal	0	2	5	7

TABLE 11

Response	Coca-cola 1	2	3	Total
drink	106	11	7	124
soda (pop)	34	16	8	58
soft drink	23	13	9	45
beverage	29	9	4	42
carbonated drink	8	17	15	32
liquid	2	11	8	21
industry (corporation)	0	3	13	16
brand name	0	8	5	13
commercial product	0	8	3	11
bottled drink	1	2	7	10

TABLE 12

Response	Collie 1	2	3	Total
dog	173	7	6	184
animal	21	35	16	72
pet	3	15	9	27
long-haired dog	0	21	5	26
mammal	1	10	10	21
canine	1	9	7	17
Lassie	0	8	4	12
large dog	0	3	7	10
friend	0	6	3	9
sheepdog	0	3	5	8

TABLE 13

Response	Congressman 1	2	3	Total
politician	47	17	3	67
government position	37	11	9	57
official	33	15	3	51
representative	19	20	11	50
elected (elected official, elected representative)	13	23	13	49
legislator	11	6	13	30
person	7	5	13	25
leader	2	13	3	18
male	2	7	5	14
crook (thief)	4	5	0	9

TABLE 14

Response	Copper 1	2	3	Total
metal	167	17	3	187
mineral	13	9	17	39
color	3	23	9	35
penny	3	21	4	28
ore	7	13	3	23
element	3	9	5	16
conductor	1	4	9	14
policeman (cop)	0	7	7	14
alloy	0	7	6	13
coin	0	7	2	9

TABLE 15

Response	Copperhead 1	2	3	Total
snake	149	23	5	177
reptile	11	31	9	51
coin	21	17	3	41
poisonous snake	1	15	19	35
dangerous (dangerous snake, dangerous animal)	0	17	11	28
animal	6	13	3	22
Indian	3	2	3	8
serpent	1	3	2	6
cold-blooded	0	3	3	6
vertebrate	0	0	6	6

TABLE 16

Response	Corporal 1	2	3	Total
Army rank	57	13	9	79
Army member	47	4	5	56
punishment	13	29	7	49
soldier	17	7	5	29
officer	17	7	1	25
body	1	14	5	20
military (member of military)	9	7	3	19
enlisted man	7	5	4	16
person	1	11	3	15
low-status	3	6	2	11

TABLE 17

Response	Daisy 1	2	3	Total
flower	193	4	0	197
plant	5	24	11	40
yellow (yellow flower, yellow object)	1	21	7	29
petaled (petaled flower; petaled object)	0	5	15	20
name	0	11	2	13
living thing	0	7	6	13
girl's name	1	8	3	12
weed	1	6	4	11
growth	0	7	3	10
fragrant (fragrant item)	0	5	4	9

TABLE 18

Response	Desk 1	2	3	Total
furniture	135	21	7	163
writing (writing table, writing surface)	21	17	3	41
table	17	15	2	34
place of study	1	17	8	26
wood product	5	8	9	22
surface (top)	5	7	4	16
school equipment	4	5	7	16
work place	4	7	3	14
office	3	7	2	12
object	2	6	3	11

TABLE 19

Response	Diamond 1	2	3	Total
stone	39	13	8	60
jewel	45	9	3	57
gem	35	14	3	52
ring	5	15	14	34
precious stone	17	5	3	25
hard (hard object)	2	14	6	22
mineral	11	3	6	20
rock	11	5	3	19
shape	2	6	7	15
jewelry	6	2	4	12

TABLE 20

Response	Dish 1	2	3	Total
utensil	73	17	2	92
plate	29	27	2	58
container	9	14	5	28
china	15	7	2	24
shape	2	11	7	20
food (meal)	3	9	3	15
woman (girl)	1	9	4	14
serving piece	7	2	4	13
pottery	2	3	3	8
fragile item	0	1	5	6

TABLE 21

Response	Door 1	2	3	Total
entrance	43	21	17	81
opening	29	17	13	59
part of house	35	11	7	53
exit	13	21	3	37
part of building	19	8	9	36
part of room	15	9	4	28
passageway	3	7	11	21
wooden (wooden thing; wooden object)	9	3	3	15
divider (partition)	2	4	5	11
barrier	2	2	4	8

TABLE 22

Response	Drum 1	2	3	Total
musical instrument	159	19	4	182
percussion instrument	17	23	10	50
sound maker (noise maker)	1	33	7	41
object to beat	5	20	7	32
round object	3	8	9	20
container	5	5	7	17
rhythm instrument	2	5	4	11
part of ear	1	7	3	11
instrument	0	6	4	10
means of communication	0	5	3	8

TABLE 23

Response	Ear			
	1	2	3	Total
body part	79	9	1	89
organ	52	7	3	62
hearihg mechanism	19	27	15	61
sensory organ (receptor organ)	16	14	8	38
part of head	11	17	2	30
part of auditory system	2	7	8	17
appendage	5	5	1	11
orifice	3	4	3	10
sound recepticle	1	6	3	10
part of face	1	3	5	9

TABLE 24

Response	Father			
	1	2	3	Total
parent	77	17	7	101
family member	59	15	9	83
relative	35	13	3	51
male (man)	11	25	11	47
authority figure	5	15	7	27
human being (person)	9	9	6	24
social role	1	5	11	17
sire	1	6	7	14
head	0	6	4	10
mother	2	4	2	8
leader	2	2	4	8

TABLE 25

Response	Fork			
	1	2	3	Total
utensil	131	17	9	157
silverware	23	15	5	43
eating utensil	19	13	2	34
pronged instrument	1	11	11	23
farm tool	1	2	4	7
split in road	0	3	3	6
shovel	0	2	4	6
division	0	2	4	6
implement	1	2	2	5
invention	0	2	3	5

TABLE 26

Response	General			
	1	2	3	Total
Army member	43	13	2	58
officer	29	16	7	52
rank	18	5	3	26
Army rank	11	8	5	24
unspecific	1	14	9	24
category	6	11	3	20
soldier	6	11	2	19
military member	6	7	4	17
specific	2	6	3	11
leader	1	7	4	11

TABLE 27

Response	Guitar			
	1	2	3	Total
musical instrument	189	6	0	195
stringed instrument	6	89	15	110
instrument	3	21	17	41
sound maker (noise maker)	0	18	7	25
wooden instrument	1	5	11	17
folk instrument	0	3	7	10
hobby	1	3	6	10
band instrument	0	2	4	6
commercial product	0	2	4	6
toy	0	2	3	5

TABLE 28

Response	Harvard			
	1	2	3	Total
school	88	13	5	106
university	55	31	8	94
college	73	12	7	92
ivy league school	9	19	7	35
institution	4	17	12	33
name	0	7	8	15
type of education	4	2	4	10
higher education institution	1	3	5	9
football team	0	1	7	8
high-status school	0	4	3	7

TABLE 29

Response	Inch 1	Inch 2	Inch 3	Inch Total
measurement (unit of measurement)	168	13	3	184
distance	5	22	7	34
type of ruler	3	16	11	30
small distance	0	20	9	29
part of foot	5	17	5	27
length	7	5	11	23
type of worm	2	6	8	16
unit	5	8	2	15
part of yard	0	2	8	10
movement	1	3	3	7

TABLE 32

Response	Mile 1	Mile 2	Mile 3	Mile Total
distance	79	39	10	128
measurement (unit of measurement)	91	17	5	113
type of race	5	30	10	45
length	12	17	6	35
long distance	1	18	9	28
unit of space	0	10	5	15
yardage	0	4	3	7
walk	1	2	3	6
traveling	0	2	3	5
abstraction	0	1	4	5

TABLE 30

Response	Lemonade 1	Lemonade 2	Lemonade 3	Lemonade Total
drink	129	7	9	145
liquid	13	25	22	60
fruit drink	8	19	11	38
soft-drink	18	9	2	29
beverage	18	8	2	28
refreshment	3	15	7	25
thirst-quencher	5	7	9	21
citrus drink	0	13	7	20
sweet drink	1	7	7	15
cold drink	2	2	6	10

TABLE 33

Response	Minute 1	Minute 2	Minute 3	Minute Total
time (unit of time)	160	11	2	173
measurement (unit of measurement)	21	13	9	43
part of hour	10	23	5	38
small amount of time	1	24	2	27
size	2	10	1	13
clock division	0	4	6	10
second	1	4	3	8
interval	1	5	2	8
short	1	3	3	7
60 sec.	0	2	3	5

TABLE 31

Response	Mayor 1	Mayor 2	Mayor 3	Mayor Total
official	70	15	3	88
politician	39	7	5	51
city official	11	25	10	46
government official	35	5	9	49
leader	2	28	11	41
elected official	11	8	19	38
head	1	21	5	27
representative	2	13	3	18
person	3	6	3	12
crook	0	6	1	7

TABLE 34

Response	Oak 1	Oak 2	Oak 3	Oak Total
tree	119	23	5	147
wood (type of wood)	52	66	13	131
plant	14	13	10	37
hard (hard wood)	1	9	9	19
living thing	2	7	7	16
leaf	1	6	3	10
tall object	0	3	6	9
type of flooring	2	3	3	8
vegetation	1	3	3	7
desk	0	1	4	5

TABLE 35

Response	Orlon			
	1	2	3	Total
synthetic	55	61	16	132
material	49	13	13	75
fabric	44	10	9	63
fiber	23	15	3	41
sweater material	1	17	6	24
cloth (type of cloth)	2	9	3	14
clothing (type of clothing)	4	5	4	13
soft material	0	5	3	8
washable material	0	1	6	7
textile	3	2	1	6

TABLE 36

Response	Parrot			
	1	2	3	Total
bird	161	11	5	177
talking bird	5	49	17	71
animal	10	23	7	40
pet	8	13	9	30
tropical bird	5	6	7	18
living thing	1	3	9	13
winged animal	3	3	4	10
repeater	0	5	5	10
multicolored	0	4	1	5
mimic	0	3	2	5

TABLE 37

Response	Pear			
	1	2	3	Total
fruit	185	6	1	192
food	9	35	8	52
shape	1	35	11	47
edible	0	21	5	26
tree	0	13	7	20
plant	0	9	10	19
sweet (sweet fruit, sweet object)	1	5	7	13
tree product	0	5	4	9
dessert	0	4	4	8
yellow (yellow object)	0	4	3	7

TABLE 38

Response	Potato			
	1	2	3	Total
vegetable	119	17	9	145
food	37	36	30	103
starch (starchy)	18	19	11	48
plant	8	11	3	22
tuber	6	8	3	17
root (root vegetable)	2	5	7	14
round object	1	11	2	14
type of form	1	3	5	9
pancake	0	3	4	7
carbohydrate	1	2	4	7

TABLE 39

Response	Scotch			
	1	2	3	Total
drink	63	14	7	84
liquor	39	29	3	71
alcohol	27	17	9	53
nationality	11	30	6	47
tape	12	16	11	39
whiskey	15	9	3	27
beverage	12	4	5	21
liquid	3	7	8	18
intoxicant	1	9	7	17
booze	2	5	3	10

TABLE 40

Response	Sedan			
	1	2	3	Total
car	133	23	3	159
automobile	39	19	5	63
vehicle	16	21	10	47
transportation (means of transportation)	1	19	13	33
type of chair (type of couch)	2	10	11	23
4-door (4-door vehicle)	0	5	3	8
2-door (2-door vehicle)	0	4	3	7
4-wheeled vehicle	0	3	4	7
machine	0	2	5	7
black vehicle	0	3	2	5

TABLE 41

Response	Shrimp 1	2	3	Total
fish	93	21	5	119
food	22	35	21	78
seafood	30	13	9	52
marine animal	10	13	13	36
shellfish	20	6	9	35
crustacean	10	13	8	31
animal	12	8	4	24
small object	1	6	4	11
size	1	4	2	7
small person	0	3	3	6

TABLE 42

Response	Sofa 1	2	3	Total
furniture	169	1	2	172
type of seat	4	47	13	64
type of bed	7	33	20	60
couch	3	22	7	32
type of chair	10	8	5	23
comfortable object	1	6	4	11
household object	1	3	4	8
living room furniture	1	3	4	8
restful place	1	3	2	6
lounge	0	0	6	6

TABLE 43

Response	Son 1	2	3	Total
family member	81	8	4	93
child	25	26	10	61
male	2	21	16	39
offspring	12	12	7	31
relative	15	7	3	25
boy	6	4	5	15
sibling	9	3	2	14
person	3	3	7	13
progeny	3	3	3	9
daughter	1	4	3	8

TABLE 44

Response	Sparrow 1	2	3	Total
bird	179	9	5	193
animal	5	28	8	41
flying (flying object)	3	24	13	40
small bird	3	13	5	21
singing bird	1	3	7	11
living thing	0	2	9	11
feathered animal	1	3	4	8
winged animal	1	5	1	7
vertebrate	1	3	3	7
mammal	1	2	3	6

TABLE 45

Response	Toe 1	2	3	Total
part of foot	47	58	5	110
part of body	79	15	11	105
part of leg	10	17	11	38
appendage	17	11	1	29
digit	18	4	2	24
extremity	9	2	5	16
has nails	0	7	6	13
phalanges	8	1	2	11
smaller member	0	4	2	6
type of dance	0	3	2	5
organ	2	1	2	5

TABLE 46

Response	Topaz 1	2	3	Total
stone	55	25	8	88
gem	50	22	4	76
jewel	45	7	9	61
color	11	8	12	31
mineral	8	14	4	26
ring	15	15	3	18
rock	6	6	2	14
semiprecious (semiprecious gem)	5	4	4	13
valuable (valuable object)	0	4	3	7
yellow (yellow object)	0	4	3	7

TABLE 47

Response	Week			
	1	2	3	Total
time (unit of time)	130	21	8	159
time period	36	15	2	53
part of month	12	23	5	40
part of year	4	25	7	36
calendar unit	12	15	3	30
composed of days	1	20	4	25
division	0	5	3	8
length	0	5	2	7
work units	0	3	2	5
measure	0	2	3	5

TABLE 48

Response	Willow			
	1	2	3	Total
tree	159	13	1	173
plant	11	38	10	59
wood	1	18	3	22
weeping	1	7	8	16
living thing	1	6	7	14
flexible	0	6	5	11
flower	4	4	2	10
branch	1	3	3	7
leaf	0	3	4	7
shady tree	0	1	4	5

TABLE 49

Response	Wool			
	1	2	3	Total
material	59	6	5	70
cloth (type of cloth)	38	18	9	65
fabric	34	9	2	45
sheep product	7	17	10	34
fiber	18	12	1	31
animal product	9	9	10	28
substance for warmth	3	8	12	23
yarn	4	6	7	17
furry	3	8	1	12
clothing material	3	2	4	9

TABLE 50

Response	Yacht			
	1	2	3	Total
boat	149	13	1	163
ship	11	13	9	33
means of transportation	6	12	10	28
status symbol	1	23	3	27
water vehicle	2	10	14	26
vessel	7	11	3	21
expensive item	1	7	11	19
sailing vessel	1	16	1	18
pleasure craft	5	6	4	15
luxury	2	10	2	12

REFERENCES

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, 1969, **80** (3, Pt. 2).

Bousfield, W. A., Cohen, B. H., & Whitmarsh, G. A. Associative clustering in the recall of words of different taxonomic frequencies of occurrence. *Psychological Reports*, 1958, **4**, 39–44.

Cohen, B. H., Bousfield, W. A., & Whitmarsh, G. A. Cultural norms for verbal items in 43 categories (ONR Tech. Rep. No. 22) Storrs: University of Connecticut, 1957.

Freedman, J. L., & Loftus, E. F. Retrieval of words from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 107–115.

Loftus, E. F., & Freedman, J. L. On predicting constrained associates from long-term memory. *Psychonomic Science*, 1970, **19**, 357–358.

Shapiro, S. I., & Palermo, D. S. Conceptual organization and class membership: Normative data for representatives of 100 categories. *Psychonomic Monograph Supplements*, 1970, **3**, 107–127.

(Received May 29, 1971)

U. S. POSTAL SERVICE

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

(Act of August 12, 1970: Section 3685. Title 39. United States Code)

SEE INSTRUCTIONS ON PAGE 2 (REVERSE)

1. TITLE OF PUBLICATION
JOURNAL OF EXPERIMENTAL PSYCHOLOGY

2. DATE OF FILING
9/27/71

3. FREQUENCY OF ISSUE
MONTHLY

4. LOCATION OF KNOWN OFFICE OF PUBLICATION *(Street, city, county, state, ZIP code) (Not printers)*
Prince & Lemon Sts., Lancaster, Lancaster, Pa. 17604

5. LOCATION OF THE HEADQUARTERS OR GENERAL BUSINESS OFFICES OF THE PUBLISHERS *(Not printers)*
1200 Seventeenth St., N.W. Washington, D. C. 20036

6. NAMES AND ADDRESSES OF PUBLISHER, EDITOR, AND MANAGING EDITOR

PUBLISHER *(Name and address)*
American Psychological Association; 1200 - 17th Street, N.W., Washington, D.C. 20036

EDITOR *(Name and address)*
David A. Grant; University of Wisconsin, Madison, Wisconsin 53706

MANAGER EDITOR *(Name and address)*
Harold P. Van Cott, 1200 - 17th Street, N.W., Washington, D.C. 20036

7. OWNER *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given.)*

NAME	ADDRESS
American Psychological Association	1200 - 17th Street, N.W., Washington, D.C. 20036

8. KNOWN BONDHOLDERS, MORTGAGEES, AND OTHER SECURITY HOLDERS OWNING OR HOLDING 1 PERCENT OR MORE OF TOTAL AMOUNT OF BONDS, MORTGAGES OR OTHER SECURITIES *(If there are none, so state)*

NAME	ADDRESS
None	

9. FOR OPTIONAL COMPLETION BY PUBLISHERS MAILING AT THE REGULAR RATES (Section 132.121, Postal Service Manual)

39 U. S. C. 3626 provides in pertinent part: "No person who would have been entitled to mail matter under former section 4359 of this title shall mail such matter at the rates provided under this subsection unless he files annually with the Postal Service a written request for permission to mail matter at such rates."

In accordance with the provisions of this statute, I hereby request permission to mail the publication named in Item 1 at the reduced postage rates presently authorized by 39 U. S. C. 3626.

(Signature and title of editor, publisher, business manager, or owner)

10. FOR COMPLETION BY NONPROFIT ORGANIZATIONS AUTHORIZED TO MAIL AT SPECIAL RATES *(Section 132.122, Postal Manual)*

(Check one)

The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes — XX Have not changed during preceding 12 months ☐ Have changed during preceding 12 months

(If changed, publisher must submit explanation of change with this statement.)

11. EXTENT AND NATURE OF CIRCULATION	AVERAGE NO. COPIES EACH ISSUE DURING PRECEDING 12 MONTHS	ACTUAL NUMBER OF COPIES OF SINGLE ISSUE PUBLISHED NEAREST TO FILING DATE
A. TOTAL NO. COPIES PRINTED *(Net Press Run)*	5,254	5,053
B. PAID CIRCULATION 1. SALES THROUGH DEALERS AND CARRIERS, STREET VENDORS AND COUNTER SALES	None	None
2. MAIL SUBSCRIPTIONS	3,929	4,065
C. TOTAL PAID CIRCULATION	3,929	4,065
D. FREE DISTRIBUTION BY MAIL, CARRIER OR OTHER MEANS 1. SAMPLES, COMPLIMENTARY, AND OTHER FREE COPIES	189	193
2. COPIES DISTRIBUTED TO NEWS AGENTS, BUT NOT SOLD	None	None
E. TOTAL DISTRIBUTION *(Sum of C and D)*	4,118	4,258
F. OFFICE USE, LEFT-OVER, UNACCOUNTED, SPOILED AFTER PRINTING	1,136	795
G. TOTAL *(Sum of E & F—should equal net press run shown in A)*	5,254	5,053

I certify that the statements made by me above are correct and complete.

(Signature of editor, publisher, business manager, or owner)
Henry S. Salmon of Bros. Mgr.

PS Form 3526 July 1971